2017

Genetic Regulation Of Tmem106b In The Pathogenesis Of Frontotemporal Lobar Degeneration

Michael Gallagher
University of Pennsylvania, mdgallagher71@gmail.com

Genetic Regulation Of Tmem106b In The Pathogenesis Of Frontotemporal Lobar Degeneration

Abstract
Neurodegenerative diseases are an emerging global health crisis, with the projected global cost of dementia alone expected to exceed $1 trillion, or >1% of world GDP, by 2018. However, there are no disease-modifying treatments for the major neurodegenerative diseases, such as Alzheimer's disease, Parkinson's disease, frontotemporal lobar degeneration (FTLD), and amyotrophic lateral sclerosis. Therefore, there is an urgent need for a better understanding of the pathophysiology underlying these diseases. While genome-wide association studies (GWAS) have identified ~200 genetic variants that are associated with risk of developing neurodegenerative disease, the biological mechanisms underlying these associations are largely unknown. This dissertation investigates the mechanisms by which common genetic variation at TMEM106B, a GWAS-identified risk locus for FTLD, influences disease risk. First, using genetic and clinical data from thirty American and European medical centers, I demonstrate that the TMEM106B locus acts as a genetic modifier of a common Mendelian form of FTLD. Second, I investigate the role of increased TMEM106B expression levels, which have been reported both in FTLD patients and in individuals carrying the TMEM106B risk allele, in FTLD pathogenesis. I demonstrate that microRNA-132, the most dysregulated microRNA in a genome-wide screen of FTLD and control brains, directly represses TMEM106B expression in human cells, and likely contributes to the elevated TMEM106B levels seen in disease. I then combine statistical approaches, bioinformatics, and experimental approaches in order to functionally characterize all candidate GWAS causal variants at the TMEM106B locus. This approach identifies a noncoding variant, rs1990620, which affects CTCF-mediated long-range chromatin interactions between distal regulatory elements, as the likely causal variant responsible for altering TMEM106B expression levels and disease risk. These results provide a plausible mechanism by which TMEM106B genotype and expression levels influence FTLD risk and clinical progression, and provide a general framework for elucidating the biological mechanisms underlying a disease-associated risk locus. Such an approach will be necessary in order to translate the thousands of loci associated with disease risk by GWAS into mechanistic understanding and therapeutic advances.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Cell & Molecular Biology

First Advisor
Alice S. Chen-Plotkin

Keywords
CTCF, eQTL, frontotemporal dementia, frontotemporal lobar degeneration, GWAS, TMEM106B

GENETIC REGULATION OF TMEM106B IN THE PATHOGENESIS OF FRONTOTEMPORAL LOBAR DEGENERATION

Michael D. Gallagher

A DISSERTATION

in

Cell and Molecular Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2017

Supervisor of Dissertation

_____

Alice S. Chen-Plotkin, M.D.
Associate Professor of Neurology

Graduate Group Chairperson

_____

Daniel S. Kessler, Ph.D.
Associate Professor of Cell and Developmental Biology

Dissertation Committee
Nancy M. Bonini, Ph.D., Florence R.C. Murray Professor of Biology
Christopher D. Brown, Ph.D., Assistant Professor of Genetics
Maja Bućan, Ph.D., Professor of Genetics (committee chair)
Stephen A. Liebhaber, M.D., Professor of Genetics
Daniel J. Rader, M.D., Seymour Gray Professor of Molecular Medicine

DEDICATION

I would like to dedicate this dissertation to my wife (Rachael), mother (Lindsay), father (Robert), and brother (Kevin), who are a constant source of love and support. I am so grateful to have such a wonderful family, without which I would not have been able to pursue all the things that I love.

ACKNOWLEDGMENTS

This dissertation is a result of the outstanding mentorship I received from Dr. Alice S. Chen-Plotkin. She has displayed incredible passion and devotion to my scientific training, and consistently provided invaluable guidance and support during my tenure in her lab.

I have also benefited from an outstanding thesis committee, consisting of Drs. Nancy M. Bonini, Christopher (Casey) D. Brown, Maja Bućan, Stephen A. Liebhaber, and Daniel J. Rader, with whom I met once per year during the dissertation phase of my thesis. Every meeting provided unique and important guidance, advice, and feedback.

I would also like to thank the faculty and administration of the Cell & Molecular Biology (CAMB) Graduate Group and the Genetics & Gene Regulation (GGR) program, for constantly taking time out of their busy schedules to help graduate students in every phase of their training. Specifically, I would like to acknowledge Drs. Daniel S. Kessler and Douglas J. Epstein, the chairpersons of CAMB and GGR at the time of my matriculation, respectively, and Meagan Schofer, Anna Kline, and Kathy O'Connor-Cooley, the CAMB administrators.

Finally, I would like to thank all past and present members of the Chen-Plotkin lab for their input, advice, and contributions.

ABSTRACT

GENETIC REGULATION OF TMEM106B IN THE PATHOGENESIS OF FRONTOTEMPORAL LOBAR DEGENERATION

Michael D. Gallagher

Alice S. Chen-Plotkin

Neurodegenerative diseases are an emerging global health crisis, with the projected global cost of dementia alone expected to exceed $1 trillion, or >1% of world GDP, by 2018. However, there are no disease-modifying treatments for the major neurodegenerative diseases, such as Alzheimer's disease, Parkinson's disease, frontotemporal lobar degeneration (FTLD), and amyotrophic lateral sclerosis. Therefore, there is an urgent need for a better understanding of the pathophysiology underlying these diseases. While genome-wide association studies (GWAS) have identified ~200 genetic variants that are associated with risk of developing neurodegenerative disease, the biological mechanisms underlying these associations are largely unknown. This dissertation investigates the mechanisms by which common genetic variation at *TMEM106B*, a GWAS-identified risk locus for FTLD, influences disease risk. First, using genetic and clinical data from thirty American and European medical centers, I demonstrate that the *TMEM106B* locus acts as a genetic modifier of a common Mendelian form of FTLD. Second, I investigate the role of increased *TMEM106B* expression levels, which have been reported both in FTLD patients and in individuals carrying the *TMEM106B* risk allele, in FTLD pathogenesis. I demonstrate that microRNA-132, the most dysregulated microRNA in a genome-wide screen of FTLD and control brains, directly represses *TMEM106B* expression in human cells, and likely contributes to the elevated TMEM106B levels seen in disease. I then combine statistical approaches, bioinformatics, and experimental approaches in order to functionally characterize all candidate GWAS causal variants at the *TMEM106B* locus. This approach identifies a noncoding variant, rs1990620, which affects CTCF-mediated long-range chromatin interactions between distal regulatory elements, as

the likely causal variant responsible for altering *TMEM106B* expression levels and disease risk. These results provide a plausible mechanism by which *TMEM106B* genotype and expression levels influence FTLD risk and clinical progression, and provide a general framework for elucidating the biological mechanisms underlying a disease-associated risk locus. Such an approach will be necessary in order to translate the thousands of loci associated with disease risk by GWAS into mechanistic understanding and therapeutic advances.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

PREFACE

This body of work aims to understand how expression levels of the gene *TMEM106B*, and the cellular regulation of these levels, are involved in the disease frontotemporal lobar degeneration (FTLD). FTLD is one of many fatal neurodegenerative diseases for which there are currently no effective treatments. Improving our understanding of what causes FTLD will help facilitate the development of therapeutic interventions. In this introduction, I will provide an overview of FTLD, discuss the role of common genetic variation in risk for complex diseases, and describe the discovery and initial characterization of *TMEM106B* as a genetic risk factor for FTLD.

CHAPTER 1: INTRODUCTION

1.1 PART I: Frontotemporal lobar degeneration: causes and mechanisms

1.1.1 Frontotemporal lobar degeneration

Frontotemporal lobar degeneration (FTLD) is a genetically and clinically heterogeneous group of diseases that are characterized by the progressive degeneration of the frontal and temporal lobes of the brain. Diagnosed clinically as frontotemporal dementia (FTD), FTLD is the second most common cause of dementia in individuals under age 65, and results in progressive deficits in behavior, executive functions, and/or language (Bang et al., 2015; Seelaar et al., 2011). In addition, up to 40% of FTLD patients experience symptoms of motor neuron disease, with 10-15% qualifying for a secondary diagnosis of amyotrophic lateral sclerosis (ALS) (Bang et al., 2015). As the disease progresses, clinical symptoms worsen, and neurodegeneration expands throughout the frontal and temporal lobes. Patients with end-stage disease display difficulties with eating, moving, and swallowing (Bang et al., 2015). Death usually occurs within ~8 years after symptom onset, and is typically caused by pneumonia or other secondary infections (Bang et al., 2015). There is currently no way to prevent, treat or cure FTLD, necessitating a better understanding of the causes and pathophysiology of this disease.

1.1.2 Epidemiology

The prevalence of FTLD is difficult to estimate due its heterogeneous clinical presentation and frequent misdiagnoses (Bang et al., 2015). In fact, diagnostic certainty usually relies on autopsy-confirmed FTLD-associated brain pathology. Various studies have estimated the prevalence of FTLD in Western countries to be roughly 10-20 per 100,000 persons in the 45-64 age group (Hodges et al., 2003; Knopman and Roberts, 2011; Lambert et al., 2014). FTLD accounts for 3-26% of dementia cases in individuals under 65, making it the second most common early-onset dementia, after Alzheimer's disease (AD) (Bang et al., 2015).

1.1.3 Neuropathology

The neuropathological features of FTLD include neuronal loss, gliosis, microvascular changes, and protein aggregation (Bang et al., 2015; Seelaar et al., 2011). Protein aggregation in

the brain is a common theme among the various neurodegenerative diseases, and different diseases are defined in part by the specific protein makeup of the aggregates seen in patient brains (Eisele et al., 2015). There are three major neuropathological subtypes of FTLD: ~50% of cases display neuronal cytoplasmic inclusions of TAR DNA-binding protein of 43kDa (TDP-43); the majority of the remaining cases exhibit various forms of neuronal and glial aggregates containing the microtubule associated protein tau (MAPT); the remaining ~5% display aggregates consisting mostly of the fused in sarcoma (FUS) protein (Bang et al., 2015; Seelaar et al., 2011). The role of these aggregates in FTLD pathogenesis is currently unknown, and is an area of intense investigation.

1.1.4 Genetic causes and risk factors

Genetics plays a complex role in FTLD. Roughly 30-50% of individuals with FTLD have a family history of dementia (familial cases), but only 10-20% of cases are transmitted in clear Mendelian inheritance patterns (Mendelian cases) (Benussi et al., 2015). Moreover, the Mendelian disease-causing mutations are not fully penetrant, and are sometimes seen in individuals with no family history of dementia (sporadic cases) (**Figures 1.1 and 1.2**) (Benussi et al., 2015; Tan et al., 2017). Roughly 40-60% of the familial cases can be attributed to mutations in the *MAPT*, *GRN*, and *C9orf72* genes, while rarer disease-causing mutations occur in the TDP-43 (*TARDBP*), valosin-containing protein (*VCP*), and charged in multivesicular body protein 2B (*CHMP2B*) genes (**Figure 1.1B**) (Bang et al., 2015; Benussi et al., 2015; Tan et al., 2017).

Whereas other common neurodegenerative diseases, such as AD and ALS, are typically sporadic with rare cases due to highly penetrant Mendelian mutations (Kalia and Lang, 2015; Kiernan et al., 2011; Scheltens et al., 2016), FTLD appears to be more genetically heterogeneous. As mentioned above, the mutations that cause Mendelian FTLD are incompletely penetrant (**Figure 1.2**), and the age at disease onset varies considerably between patients (Benussi et al., 2015). Thus, it is likely that other genetic and environmental factors modify the penetrance and clinical manifestation associated with these mutations.

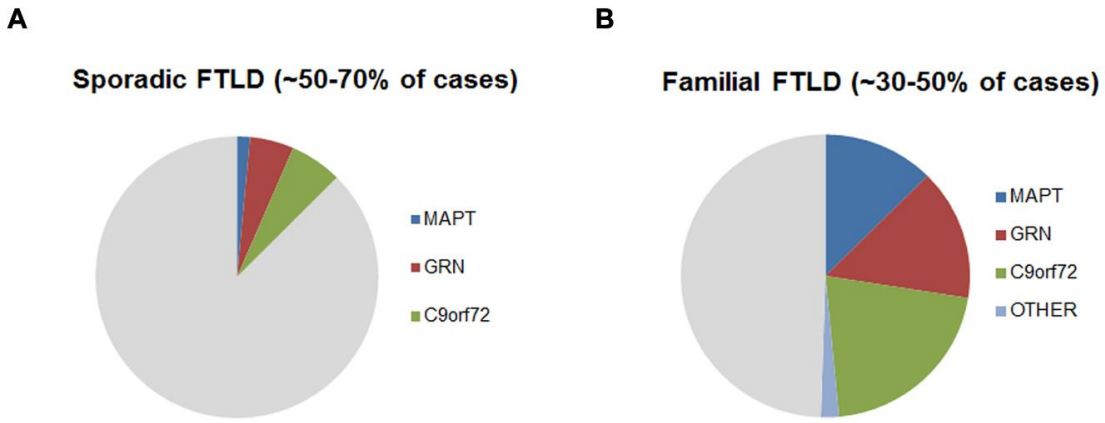


Figure 1.1. The prevalence of FTLD-causing mutations.
The prevalence of FTLD-causing mutations in **(A)** sporadic and **(B)** familial FTLD cases. "Other" mutations occur in the *VCP*, *CHMP2B*, and *TARDBP* genes. Gray slices of the pie charts indicate sporadic **(A)** and familial **(B)** cases that do not have any of the indicated mutations. Charts are based on data from Benussi et al., 2015, *Frontiers in Aging Neuroscience*.

In addition to disease-causing mutations, two genome-wide association studies (GWAS) have been performed to identify genetic risk factors for FTLD. A GWAS conducted at the University of Pennsylvania was performed using only FTLD cases with TDP-43 pathology (FTLD-TDP), and a genome-wide significant association was found on chromosome 7p21, at the *TMEM106B* locus (Van Deerlin et al., 2010). This association has been independently replicated (Finch et al., 2011; Hernandez et al., 2015; van der Zee et al., 2011). A second GWAS was performed using a clinically defined FTLD cohort (thus likely representing various neuropathological forms of FTLD), and a significant association with risk for FTLD was found at chromosome 6p21.3, at the *HLA* locus (Ferrari et al., 2014). The lack of association with genetic variation at the *TMEM106B* locus in the clinical FTLD GWAS suggests that distinct disease mechanisms may underlie neuropathologically-defined FTLD, versus clinically-defined FTLD that results from multiple potential neuropathologies.

1.1.4.1 *MAPT*

The first Mendelian FTLD-causing mutations were found in the *MAPT* gene (Hutton et al., 1998; Poorkaj et al., 1998; Spillantini et al., 1998). *MAPT* encodes the microtubule associated protein tau (MAPT), which promotes microtubule assembly and stability. MAPT is highly expressed in neurons and localizes mainly to axonal processes, although it is expressed at lower levels in glial cells as well (Mackenzie and Neumann, 2016). In the human brain, alternative splicing of *MAPT*

produces six protein isoforms, three of which contain three amino acid repeats (3R tau), and three of which contain four amino acid repeats (4R tau) (Mackenzie and Neumann, 2016). More than 50 autosomal dominant mutations in *MAPT* have been discovered in FTLD patients, comprising ~10% of Mendelian FTLD cases (Bodea et al., 2016; Mackenzie and Neumann, 2016). The two major classes of mutations are 1) missense or deletion mutations in exons 9-13 that impair the microtubule-stabilizing function of MAPT and 2) intronic and coding mutations that affect the splicing of exon 10, resulting in alterations of the 3R:4R MAPT ratios (Mackenzie and Neumann, 2016). These mutations have been detected in a small number (0-3%) of sporadic FTLD cases, suggesting either incomplete penetrance or *de novo* mutation (Benussi et al., 2015) (**Figure 1.1A**).

All FTLD cases caused by *MAPT* mutations display MAPT-associated neuropathology, and are thus referred to as "FTLD-tau". Most FTLD-tau cases, however, are not caused by *MAPT* mutations, implicating other mechanisms in MAPT-associated neuropathology in FTLD (Bodea et al., 2016). In FTLD-tau, abnormal intracellular inclusions of hyperphosphorylated MAPT occur in neurons, and in some cases, glia (Bodea et al., 2016). These inclusions result in the mislocalization of MAPT from the axons to the cell body (soma) and dendrites. Interestingly, other forms of MAPT aggregates, called neurofibrillary tangles, are one of the hallmark neuropathological features of AD (Bodea et al., 2016).

The mechanisms by which *MAPT* mutations and aggregates cause FTLD and other neurodegenerative diseases are unknown. Normal phosphorylation of MAPT may be required for normal interactions with microtubules, whereas hyperphosphorylation, associated with disease states, sequesters MAPT from axons and results in conformational changes and eventual aggregation (Bodea et al., 2016). While various *MAPT* knockout animal models have been created, the existence of compensatory genes in some lower organisms has made it difficult to draw conclusions regarding *MAPT* function (Bodea et al., 2016). Consequently, much of our understanding of the role of *MAPT* in neurodegeneration has come from transgenic rodent models. While expression of human wild-type (WT) MAPT in rodents causes only mild phenotypes, including MAPT hyperphosphorylation and mislocalization, expression of FTLD-associated *MAPT*

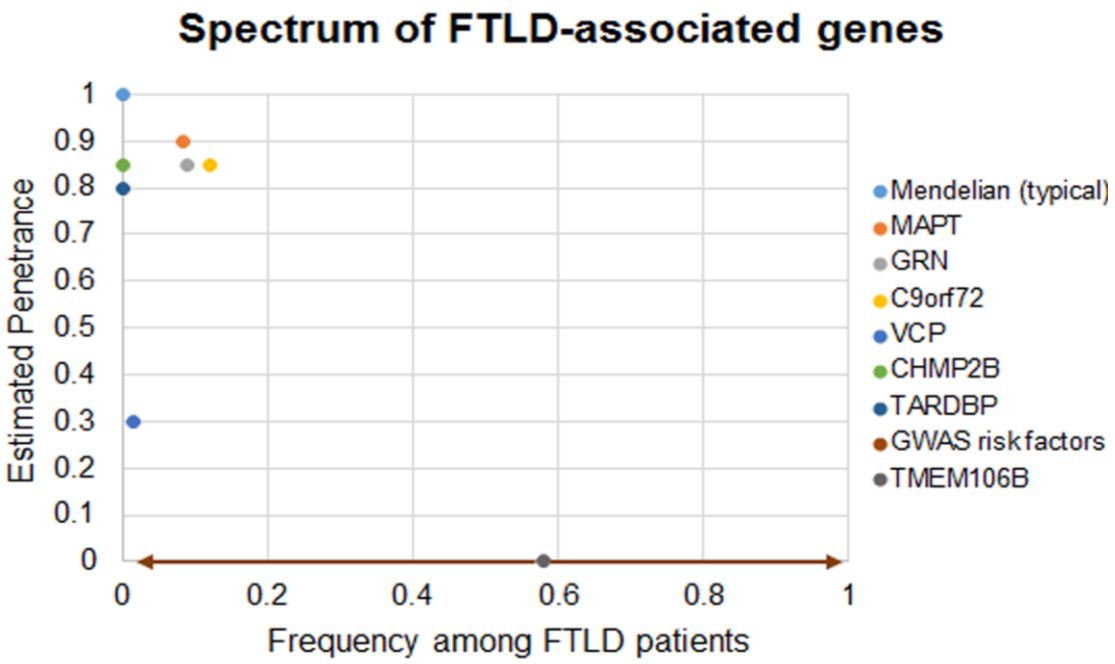


Figure 1.2. Estimated penetrance and frequencies of FTLD-associated mutations and risk variants.
The estimated penetrance and frequencies among FTLD patients is plotted for all known FTLD-causing mutations, as well as a typical, fully penetrant Mendelian mutation, typical GWAS-identified risk factors, and *TMEM106B* risk variants. Values on the x-axis represent percentages (e.g. 0.2=20%), whereas values on the y-axis are relative (0=never causes disease, 1=always causes diseases). Penetrance estimates are based on Benussi et al., 2015, *Frontiers in Aging Neuroscience*. Since risk factors do not cause disease, but rather affect risk, their estimated penetrance is zero. Also note that while GWAS risk variants typically have allele frequencies of at least ~1% in the general population, they can theoretically be much more common due to lack of negative selection pressures and other factors.

mutations causes neurofibrillary tangles, neuronal loss, gliosis, and motor symptoms (Bodea et al., 2016). However, the molecular mechanisms and cellular pathways by which this occurs in animal models and human disease are unknown.

1.1.4.2 *GRN*

The second Mendelian cause of FTLD to be discovered was mutations in the *GRN* gene, which encodes progranulin (PGRN), a secreted growth factor that functions in inflammation, wound repair, neuronal viability, and cancer. PGRN is expressed in many cell types, including neurons and glia, and is upregulated in activated microglia and astrocytes (Petkau and Leavitt, 2014). When PGRN is secreted, extracellular proteases cleave the protein into smaller granulin peptides. Full-length PGRN and cleaved granulins appear to have opposing and complex roles in

neuroinflammation (Mackenzie and Neumann, 2016; Petkau and Leavitt, 2014). PGRN also facilitates neurite outgrowth, neurogenesis, and neuronal survival, and neurons are at least partially dependent upon PGRN secretion by glial cells (Martens et al., 2012; Petkau and Leavitt, 2014; Raitano et al., 2015). At least 69 loss-of-function mutations in *GRN* have been discovered, accounting for 10-20% of familial FTLD cases (Cruts et al., 2006; Gass et al., 2006) (**Figure 1.1B**), and all result in PGRN haploinsufficiency (Rademakers et al., 2012). Most disease-causing mutations create a premature stop codon, resulting in nonsense-mediated decay of the mature transcript. Other mutations impair the trafficking and eventual secretion of PGRN, highlighting the important role of PGRN secretion in normal brain function (Kleinberger et al., 2013; Petkau and Leavitt, 2014). As with *MAPT* mutations, all disease-causing *GRN* mutations display autosomal dominant inheritance patterns (Petkau and Leavitt, 2014). Interestingly, a functional single nucleotide polymorphism (SNP) in the *GRN* 3' untranslated region, rs5848, alters *GRN* mRNA levels by disrupting a binding site for the microRNA miR-659. The allele of rs5848 that preserves the miR-659 binding site results in reduced PGRN levels, and has been reported to increase risk for FTLD-TDP (Rademakers et al., 2008). However, attempts to replicate this association have produced conflicting results (Chen et al., 2015; Rollinson et al., 2011b; Simon-Sanchez et al., 2009).

All FTLD cases caused by *GRN* mutations (*GRN*+ FTLD) display TDP-43 neuropathology (FTLD-TDP) (Petkau and Leavitt, 2014). TDP-43, which is normally a predominantly nuclear protein, mislocalizes to the cytoplasm in FTLD-TDP, becomes hyperphosphorylated and ubiquitinated, and forms intraneuronal aggregates (Ratti and Buratti, 2016). These aggregates are thought to contribute to disease by impairing the normal function of TDP-43 and/or conferring a toxic gain-of-function (Ratti and Buratti, 2016). In addition, FTLD-TDP brains display severe gliosis in affected brain regions. PGRN is drastically upregulated in these regions, so much so that *GRN*+ FTLD-TDP individuals have higher PGRN levels in the frontal cortex than healthy individuals, who have two functional copies of the *GRN* gene (Chen-Plotkin et al., 2010b). Thus, PGRN haploinsufficiency may cause disease in part by disrupting the neuroinflammatory response,

although it is difficult to determine whether this phenotype is a cause of disease, a result of disease, or both. As mentioned above, *GRN* mutations are incompletely penetrant (**Figure 1.2**), and FTLD symptoms can manifest anywhere from ~35 to ~89 years of age (Benussi et al., 2015) in mutation carriers. These observations suggest that other genetic and environmental factors influence *GRN*+ FTLD-TDP manifestation.

Several animal models of *GRN*+ FTLD-TDP have been created. Surprisingly, heterozygosity of *Grn* in mice (which causes FTLD-TDP in humans) produces no obvious phenotype, despite 75% amino acid similarity between the human and mouse orthologues. *Grn* knockout mice do display some FTLD-relevant phenotypes, such as gliosis, ubiquitinated (but usually TDP-43-negative) neuronal inclusions, and subtle behavioral phenotypes, but do not develop overt neurodegeneration (Filiano et al., 2013; Ma et al., 2016; Martens et al., 2012; Petkau and Leavitt, 2014; Tanaka et al., 2013a; Tanaka et al., 2013b; Wils et al., 2012; Yin et al., 2010a; Yin et al., 2010b). Paradoxically, in humans, homozygous *GRN* null individuals develop an entirely different syndrome, neuronal ceroid lipofuscinosis, which is a fatal lysosomal storage disease (Smith et al., 2012). Thus, animal models of *GRN*+ FTLD-TDP may be of limited utility, and have not elucidated the mechanisms underlying the human disease, although they do suggest a potential lysosomal role for *GRN* in disease.

Since *GRN*+ FTLD-TDP results from PGRN haploinsufficiency, cell-based studies have also been carried out to investigate the effects of the loss of PGRN on neuronal health. One study in particular demonstrated that exposing rat neurons to conditioned medium from *Grn* knockout microglia caused neuronal death, suggesting a role for microglial PGRN in maintaining neuronal health (Martens et al., 2012). Other reports have further implicated PGRN in microglial function, including endocytosis and *in vivo* synaptic pruning (Lui et al., 2016; Pickford et al., 2011). PGRN deficiency also appears to result in altered lysosomal function (thus linking loss of PGRN to the lysosomal storage diseases seen in PGRN null humans) and an increase in Wnt signaling pathways (Alquezar et al., 2016; de la Encarnacion et al., 2016; Rosen et al., 2011; Tanaka et al., 2017).

Finally, *GRN* has also been implicated in other neurodegenerative diseases. Specifically, *GRN* overexpression and knockdown have been shown to alleviate and enhance the toxic effects of amyloid beta protein, respectively, in an AD mouse model (Minami et al., 2014). Moreover, the rs5848 *GRN* variant, which has been reported to affect risk for FTLD-TDP, may also affect risk for AD (Lee et al., 2011; Sheng et al., 2014), although these results require replication.

In summary, *GRN* is a common Mendelian cause of FTLD-TDP, accounting for 10-20% of all FTLD-TDP cases (Gass et al., 2006). All FTLD-causing *GRN* mutations act through a loss-of-function mechanism, implicating restoration of PGRN levels as a potential therapeutic target. *GRN* has complex roles in inflammation, wound repair, neuronal viability, autophagy, and other processes, and its effects appear to act in both cell autonomous and non-cell autonomous manners (Petkau and Leavitt, 2014). Improving our understanding of the role of *GRN* in FTLD is an important basic and translational research goal.

1.1.4.3 *C9orf72*

In 2011, hexanucleotide repeat expansion (HRE) mutations in the *C9orf72* gene were discovered to be the most common cause of FTLD, ALS, and the combined syndrome FTLD/ALS (DeJesus-Hernandez et al., 2011; Renton et al., 2011). While most normal individuals have 2-20 GGGGCC (G_4C_2) repeats at the 5' end of the *C9orf72* gene (corresponding to the promoter region for one transcript variant and the first intron for the other two transcript variants of this gene), individuals with disease have hundreds to thousands of repeats, and these HREs cause disease in an autosomal dominant manner (DeJesus-Hernandez et al., 2011; Renton et al., 2011; Todd and Petrucelli, 2016). Similar to *GRN* and *MAPT* mutations, penetrance is incomplete (**Figure 1.2**), and age at disease onset varies considerably (Benussi et al., 2015). While the normal function of *C9orf72* is not known, several studies have suggested a role in endosomal trafficking and autophagy (Farg et al., 2014; Sellier et al., 2016; Sullivan et al., 2016; Ugolino et al., 2016; Webster et al., 2016; Yang et al., 2016).

Both FTLD and ALS cases caused by *C9orf72* HREs display TDP-43 pathology (*C9orf72*+ FTLD-TDP), similar to *GRN*+ FTLD-TDP and most sporadic ALS cases (Al-Sarraj et al., 2011;

Neumann et al., 2006). Thus, FTLD and ALS display neuropathological overlap, in addition to genetic and clinical overlap. *C9orf72*+ FTLD-TDP cases also display inclusions negative for TDP-43 but positive for Nucleoporin 2 (p62), a protein complex associated with the nuclear envelope, throughout the central nervous system (Al-Sarraj et al., 2011). The mechanism by which the *C9orf72* HRE causes disease has been an area of intense investigation in recent years. Many studies have supported toxic-gain-of-function mechanisms, but loss of normal *C9orf72* function may also play a role in disease. I briefly summarize these studies below.

The presence of hundreds to thousands of G_4C_2 repeats in the *C9orf72* promoter and first intron (depending on the transcript variant) has been proposed to negatively affect *C9orf72* transcription (DeJesus-Hernandez et al., 2011). In support of this, studies have reported reduced *C9orf72* transcript and protein levels in patient cells and brain tissue (Ciura et al., 2013; DeJesus-Hernandez et al., 2011; Donnelly et al., 2013; Gijselinck et al., 2012; Tran et al., 2015). However, the effects of these reduced levels in patients are unknown, as *C9orf72* knockdown and knockout have produced conflicting results in animal models. For example, reduction of *C9orf72* levels in both *C. elegans* and zebrafish result in neuronal and behavioral defects (Ciura et al., 2013; Therrien et al., 2013); however, similar knockdown experiments in mice did not produce similar phenotypes. *C9orf72* knockout mouse models have reproducibly displayed altered immune responses, (specifically, enlarged spleen and lymph nodes, lysosomal accumulation, and neuroinflammation), but no neurodegeneration or related phenotypes (Atanasio et al., 2016; Jiang et al., 2016; Koppers et al., 2015; Lagier-Tourenne et al., 2013; O'Rourke et al., 2016; Sudria-Lopez et al., 2016). Finally, a patient homozygous for the *C9orf72* HRE was reported to have even lower *C9orf72* transcript levels than those seen in heterozygotes, but had similar clinical and neuropathological features (Fratta et al., 2013). While additional HRE homozygotes are needed to establish whether this is a reproducible pattern, these results suggest that loss-of-normal function of *C9orf72* may not play a significant role in FTLD-TDP pathogenesis.

In support of the toxic-gain-of-function model, various animal models have been generated that express HREs of various lengths, usually >30 repeats to mimic disease and <10 repeats to

mimic healthy individuals. For example, expression of 30 repeats in the *Drosophila* eye or motor neurons results in neurodegeneration (Xu et al., 2013), and expression of 66 repeats in mouse neurons results in TDP-43 inclusions in the brain, neurodegeneration, and behavioral deficits (Chew et al., 2015). While some bacterial artificial chromosome (BAC) transgenic mouse models expressing human *C9orf72* with repeat expansions do not display TDP-43 pathology or neurodegeneration (O'Rourke et al., 2015; Peters et al., 2015), others display neurodegeneration, behavioral deficits, and in some cases, decreased lifespan (Jiang et al., 2016; Liu et al., 2016). Since all of these studies use BACs containing similar genomic regions, the reason for the discrepant results is unclear but raises the possibility that off-target effects might be causing some of the phenotypes.

Various toxic-gain-of-function mechanisms for the *C9orf72* HRE have been proposed, and in many cases, validated. At the biochemical level, the HRE forms DNA and RNA G-quadruplexes as a result of its G-rich sequence. One study demonstrated that nucleolin, a key nucleolar protein, is sequestered by *C9orf72* HRE RNA quadruplexes (Haeusler et al., 2014). Nucleolin localization and nucleolar morphology is disrupted in patient cells and brain tissue, suggesting a role of nucleolar stress in disease (Haeusler et al., 2014). In addition, HRE-expressing animals, patient cells (typically fibroblasts from *C9orf72*+ FTLD-TDP patients that are differentiated into induced pluripotent stem cell-derived neurons), and patient brains consistently display RNA foci consisting of the transcribed HRE sequence (DeJesus-Hernandez et al., 2011; Donnelly et al., 2013; Gendron et al., 2013; Lagier-Tourenne et al., 2013; Mizielinska et al., 2013; Renton et al., 2011; Zu et al., 2013). Many studies have demonstrated that these foci sequester RNA-binding proteins (RBPs), resulting in gene expression changes, as well as sensitivity to glutamate toxicity, although the specific affected RBPs differ widely across studies (Cooper-Knock et al., 2014; Donnelly et al., 2013; Haeusler et al., 2014; Lee et al., 2013; Mori et al., 2013; Sareen et al., 2013). However, siRNA or shRNA-mediated knockdown of these foci have rescued cellular and organismal phenotypes induced by the HRE, suggesting that they (or the dipeptide repeat proteins also

produced from the HRE, see below) are toxic (Donnelly et al., 2013; Jiang et al., 2016; Sareen et al., 2013).

The HRE transcript also produces dipeptide protein repeats (DPRs) through repeat-associated non-ATG (RAN) translation (Zu et al., 2011), which are seen in patient cells, patient brain, and some animal models (Al-Sarraj et al., 2011; DeJesus-Hernandez et al., 2011; Mann et al., 2013; Murray et al., 2011; Renton et al., 2011). Several studies have shown that these DPRs are toxic when expressed in yeast, cultured neurons, *Drosophila*, and mice (Kwon et al., 2014; Mizielinska et al., 2014; Tao et al., 2015; Wen et al., 2014; Yang et al., 2015; Zhang et al., 2016). The pathogenicity of DPRs is further supported by genetic modifier screens performed in *Drosophila* and yeast. These studies, using HRE models that produce RNA foci, DPRs, or both, have identified proteins involved in nucleocytoplasmic transport as genetic modifiers of HRE toxicity (Freibaum et al., 2015; Jovicic et al., 2015; Zhang et al., 2015). In general, the results suggest that the HRE impairs the nuclear import, and possibly export, of mRNAs, resulting in the accumulation of mRNAs in the nucleus (Freibaum et al., 2015; Jovicic et al., 2015; Zhang et al., 2015).

Finally, a toxic gain-of-function mechanism is also supported by the disease-modifying effects of CpG methylation in the *C9orf72* promoter. Hypermethylation of the promoter has been shown to reduce *C9orf72* levels and resulting RNA foci and DPRs, as well as protect against cellular sensitivity to stress (Liu et al., 2014; Russ et al., 2015). Importantly, this hypermethylation also appears to reduce neurodegeneration, delay age at death, and extend disease duration in human patients (McMillan et al., 2015).

In summary, hexanucleotide repeat expansions in the *C9orf72* gene are the most common cause of FTLD, as well as ALS. *C9orf72* appears to be involved in autophagy, although the function of this protein is largely unknown. The HREs likely cause disease through a toxic gain-of-function mechanism (driven by toxic RNA foci and DPRs), although loss of normal *C9orf72* function may play a role as well. The normal and disease-associated functions of *C9orf72* are currently an area of intense investigation.

1.1.4.5 Rare mutations

As mentioned above, ~50% of familial FTLD cases are caused by mutations in the *MAPT*, *GRN* and *C9orf72* genes. These three genes appear to have quite different cellular functions, although *GRN* and *C9orf72*, which both cause FTLD-TDP, have been implicated in autophagy, suggesting that protein degradation pathways may play a role in FTLD pathogenesis. Intriguingly, rarer FTLD-causing mutations also point to defects in autophagy, as well as the ubiquitin/proteasome system, the other major protein clearance pathway in the cell (Cohen-Kaplan et al., 2016). This genetic evidence that impaired protein degradation may play a role in FTLD is strengthened by the fact that such impairments have been implicated in virtually all neurodegenerative diseases (Martini-Stoica et al., 2016).

Some rare FTLD cases are caused by autosomal dominant mutations in the *CHMP2B* gene, which is part of the endosomal sorting complex required for transport-III (ESCRT-III) complex (Isaacs et al., 2011) (**Figure 1.1B**). This complex plays a role in protein degradation through the endolysosomal and autophagic pathways. Two FTLD-causing mutations, one affecting splicing and another introducing a premature stop codon, both result in a truncated protein lacking the C-terminal 36 or 49 amino acids, respectively (Isaacs et al., 2011). The mechanism by which the mutation causes disease is unknown, but multiple studies have demonstrated an accumulation of autophagic vesicles in cells expressing *CHMP2B* with FTLD-causing mutations (Isaacs et al., 2011). In addition, mice expressing an FTLD mutant form of *CHMP2B* develop a lysosomal storage disease (Clayton et al., 2015). Patients with these mutations have ubiquitin/p62-positive inclusions, thus referred to as FTLD-UPS.

Other rare FTLD cases are caused by various missense mutations in the *VCP* gene (Kimonis et al., 2008) (**Figure 1.1B**). VCP is an (AAA) ATPase and plays a critical role in the ubiquitin/proteasome protein degradation pathway. The disease-causing mutations cause a distinct syndrome called inclusion body myopathy associated with Paget disease of the bone and frontotemporal dementia (IBMPFD), which causes muscle weakness and bone disease in addition

to FTLD (Mackenzie and Neumann, 2016). IBMPFD is characterized by TDP-43 pathology, and is also inherited in an autosomal dominant manner (Erzurumlu et al., 2013; Kimonis et al., 2008).

Finally, rare mutations in *TARDBP*, the gene encoding TDP-43, have been reported to cause ALS, and in even rarer cases, FTLD (**Figure 1.1B**), although the FTLD cases typically lack neuropathological confirmation (Mackenzie and Neumann, 2016). Nonetheless, the important role of TDP-43 in FTLD pathogenesis has been appreciated since 2006, when TDP-43 was reported to be the main protein constituent of the neuronal cytoplasmic ubiquitinated inclusions seen in ~50% of FTLD brains (originally referred to as FTLD-U, now referred to as FTLD-TDP) (Neumann et al., 2006). TDP-43 is a heterogeneous ribonucleoprotein (hnRNP) that has widespread roles in RNA processing, regulation, splicing, and stability (Ratti and Buratti, 2016). Experimental knockdown of TDP-43 in mouse and human cells results in changes in expression levels and splicing patterns of thousands of mRNAs (Polymenidou et al., 2011; Tollervey et al., 2011), and may also result in widespread changes in other RNAs, such as microRNAs and long non coding RNAs (lncRNAs). TDP-43 contains N- and C-terminal domains, which are thought to play a role in protein-protein interactions, and two RNA recognition motifs (RRMs), which provide sequence specific RNA-binding properties (Ratti and Buratti, 2016). Over 50 disease-causing TDP-43 mutations have been reported (the majority of which cause ALS, not FTLD), and occur mostly in the C-terminal region, suggesting that they affect the interactions of TDP-43 with itself and/or other proteins. Many of these mutations cause neuropathology or neurodegeneration when expressed in animal models, and some enhance the ability of TDP-43 to form aggregates both *in vitro* and *in vivo* (Ratti and Buratti, 2016). In general, however, TDP-43 animal models do not fully recapitulate the phenotypes associated with ALS or FTLD, and the effects of the majority of disease-associated TDP-43 mutations are either unknown or contradicted in the literature (Buratti E., 2015).

In summary, known FTLD-causing mutations occur in six genes (**Figure 1.1B**). All but MAPT have been implicated in protein degradation pathways; indeed, the primary cellular functions of *CHMP2B*, *VCP*, and possibly *C9orf72* may be directly related to protein degradation. *TARDBP* has widespread roles in RNA processing, splicing, and metabolism, and thus likely plays a global

(albeit indirect) role in regulating protein levels. Finally, accumulating evidence from cell-based studies and animal models also implicate *GRN* in autophagy. Taken together, this genetic evidence suggests that defects in protein homeostasis may play an important role in FTLD, making autophagy and other processes related to protein degradation attractive therapeutic targets.

1.1.4.6 *TMEM106B*

While much progress has been made in understanding the genetic basis of FTLD, much work remains to be done. As mentioned previously, known mutations account for only ~50% of familial FTLD cases, and the mechanisms by which these mutations cause disease are not well understood. Secondly, 80-90% of FTLD cases are not inherited in Mendelian patterns (Benussi et al., 2015), suggesting that they are caused by combinations of genetic and environmental risk factors – in other words, FTLD is a complex trait. As is the case for many complex traits, researchers have used the GWAS to identify genetic risk factors for FTLD, with the goal of expanding our understanding of the pathways involved in disease, in order to help facilitate the development of therapeutic interventions.

The first FTLD GWAS was performed on 515 FTLD-TDP cases and 2,509 controls of Western European ancestry (Van Deerlin et al., 2010). Three SNPs in moderate to complete linkage disequilibrium (LD) with each other ($0.7 \leq r^2 \leq 1$) were significantly associated with risk for FTLD-TDP, with the major (more common) SNP alleles associating with increased disease risk ($P=1.08 \times 10^{-11}$ and odds ratio=1.64 for the top SNP, rs1990622). The SNPs are located on chromosome 7p21; two are located in intronic regions of the previously uncharacterized gene transmembrane protein 106B (*TMEM106B*), and the third, most significant (sentinel) SNP is located ~7kb downstream of *TMEM106B* (Van Deerlin et al., 2010). Importantly, the association of rs1990622 with FTLD was replicated both in the initial GWAS report, as well as in three independent studies (Finch et al., 2011; Hernandez et al., 2015; van der Zee et al., 2011).

Supporting a more general role of *TMEM106B* genotype in neurodegeneration, the major (risk) allele of rs1990622 and other linked SNPs have also been associated with hippocampal sclerosis of aging (Nelson et al., 2015), cognitive impairment in ALS (Vass et al., 2011), increased

TDP-43 pathology in Alzheimer's disease and neurologically normal individuals (Rutherford et al., 2012; Yu et al., 2015), decreased neuronal connectivity in asymptomatic *GRN* mutation carriers (Premi et al., 2014), reduced plasma progranulin levels (Cruchaga et al., 2011; Finch et al., 2011), and reduced temporal lobe gray matter and smaller anterior commissure cross-sectional area in neurologically healthy individuals (Adams et al., 2014). While I discuss individual studies in greater detail in Chapter 1.3, these results suggest that common genetic variation at the *TMEM106B* locus affects risk for FTLD, as well as several other phenotypes associated with various neurodegenerative diseases.

1.1.5 Summary

FTLD refers to a heterogeneous group of neurodegenerative diseases characterized by progressive degeneration of the frontal and temporal lobes, behavioral changes, language problems, executive function deficits, and oftentimes, motor neuron disease. Up until ~10 years ago, there was very little understanding of the genetic causes, pathophysiology, and neuropathology associated with FTLD; as a result, FTLD was (and still is) frequently misdiagnosed as other types of dementia, psychiatric illness, and other conditions. The recent consolidation of findings from the fields of genetics, neurology and neuropathology has refined our understanding of the FTLD disease spectrum, and FTLD is now known to be a major neurodegenerative disease, accounting for up to a quarter of all pre-senile dementia cases (Bang et al., 2015).

FTLD is a relatively heritable disease, with cases frequently presenting with a family history of disease, making it unique among neurodegenerative diseases in general. The most common disease-causing mutations are loss-of-function mutations in the *GRN* and *MAPT* genes and hexanucleotide repeat expansion mutations in the *C9orf72* gene, although these mutations are not fully penetrant. In addition, most FTLD cases are of unknown cause, and the genes that harbor disease-causing mutations have diverse and somewhat non-overlapping cellular roles. Thus, our understanding of the mechanisms underlying FTLD is limited, necessitating further advances in basic research that can be applied to nominate potential therapeutic targets. This is an especially

important research goal, given that there are currently no disease-modifying treatments for this fatal disease.

One approach that has been used to identify novel pathways and mechanisms involved in FTLD is the GWAS, which has identified thousands of risk loci for hundreds of complex traits and diseases (Welter et al., 2014). The first GWAS performed on FTLD identified an association of SNPs on chromosome 7p21 with FTLD risk. Interestingly, these SNPs fall in and near the previously uncharacterized gene *TMEM106B* (Van Deerlin et al., 2010). However, the statistical association alone does not demonstrate that *TMEM106B* itself is the causal gene, nor how genetic variants at this locus might affect *TMEM106B* and/or other genes to influence disease risk. Thus, it will be critical to identify how common genetic variation at this locus influences risk for FTLD.

1.2 PART II: The role of common genetic variation in complex traits

1.2.1 Complex traits and genome-wide association studies

For many decades after the discovery of the structure of DNA and the genetic code, the field of human genetics was largely focused on understanding the structure and function of protein-coding genes, and how mutations in these genes cause disease. The Central Dogma of molecular biology posits that genes are first transcribed into messenger RNA (mRNA), after which the mRNA is translated into protein (Crick, 1970). Because of the seemingly universal nature of the genetic code, it was easy to predict how simple or complex alterations of the underlying DNA sequence would change the amino acid composition of the resulting protein (Ghosh and Collins, 1996). In addition, it was clear from Mendelian genetics that diseases that run in families in predictable patterns are caused by mutations in a single gene. In the latter part of the 20th century, the causative mutations underlying many Mendelian diseases were elucidated by positional cloning (Ghosh and Collins, 1996), and thus, an important hurdle was overcome in our understanding of the genetic bases of human disease.

Today, the genetic lesions responsible for most of the common (and many of the rare) Mendelian diseases are known, and in many cases researchers have determined how the mutation in question affects protein function and human physiology (Ghosh and Collins, 1996). However, many of the most common and deadly diseases, such as cardiovascular disease, cancer, Alzheimer's disease, Parkinson's disease, and Type 2 diabetes, are typically not (or never) caused by single mutations (Hirschhorn, 2005; Johnson and Todd, 2000). Such "complex traits" are instead influenced by a combination of multiple genetic and environmental risk factors, and thus do not follow Mendelian inheritance patterns (Ghosh and Collins, 1996). This situation poses a formidable challenge to scientists trying to understand the causes of these diseases. While it may not always be necessary to understand the cause of a disease in order to successfully treat it, such a mechanistic understanding certainly increases the likelihood that a successful therapeutic intervention will be achieved.

For several reasons, elucidation of the genetic basis of a complex trait is a more daunting task than identifying a disease-causing mutation (Ghosh and Collins, 1996). Family pedigrees are of little use, and complex traits, by definition, are influenced by many genes (human height, for example, appears to be affected by genetic variation at several hundred loci across the genome (Wood et al., 2014)). Starting in 2005, researchers around the world have used the genome-wide association study (GWAS) to identify loci that harbor genetic variants (typically single nucleotide polymorphisms, or "SNPs") that are associated with risk for complex diseases and traits (Edwards et al., 2013). The GWAS era has been successful in the sense that thousands of loci have been statistically associated with risk for hundreds of diseases and traits, and a notable number of these loci are well-replicated, suggesting that they are true associations (Welter et al., 2014). However, several factors have made it difficult to bridge the gap between statistical associations and a functional understanding of how these variants influence disease risk. First, the association of a locus with disease does not specify which variant at that locus is actually *causing* the association (the "causal variant"), nor which gene is affected by the causal variant. The former problem is due to the fact that there are often many co-inherited variants in strong linkage disequilibrium (LD) with the most significant (sentinel) disease-associated variant, comprising a "haplotype" (Gabriel et al., 2002); within the haplotype, SNPs and other genetic variants in strong LD often have statistically indistinguishable associations with disease risk. Therefore, only experimental validation can determine which of the linked variants are functional (Edwards et al., 2013; Schaub et al., 2012). The latter complication results from the fact that >90% of disease-associated variants (daVs) are located in non-protein-coding regions of the genome, and many are far away from the nearest known gene (Maurano et al., 2012; Schaub et al., 2012). What might these non-coding variants be doing? One clue arises from the observation that daVs, as well as variants in strong LD with them, are enriched in predicted transcriptional regulatory regions, called "*cis*-regulatory elements" (CREs) (Maurano et al., 2012; Schaub et al., 2012). This suggests that many loci implicated by GWAS to affect disease risk may do so by altering the genetic regulation of one or more target genes. However, the complex nature of eukaryotic transcriptional regulation (Lelli et al., 2012) can

make it difficult to assign putative CREs (and any disease-associated variants within them) to their correct target genes (Edwards et al., 2013), necessitating the use of genomic data sets and experimental approaches to help answer this question. Indeed, while several thousand GWAS have been performed, and many thousands of loci have been confirmed as bona fide disease risk factors (Welter et al., 2014), the number of studies that have investigated the mechanisms underlying particular associations number only in the dozens (**Figure 1.3**), and the number of studies that have functionally characterized all candidate causal variants at a given locus are even fewer (**Table 1.1** lists some recent examples, but is by no means exhaustive). The purpose of this review is to present a general framework for the functional dissection of a disease-associated risk locus, and to highlight individual studies as proof-of-principle examples for the various approaches that have been used by researchers in mechanistic GWAS follow-up studies.

1.2.2 The role of gene expression in complex traits

As mentioned above, the vast majority of daVs reside in noncoding regions of the genome, suggesting that these variants may affect gene expression through effects on transcription, splicing, or mRNA stability. Consistently, several studies have shown that daVs are enriched in predicted CREs, typically defined by chromatin accessibility (as determined by DNase-seq, FAIRE-seq, ATAC-seq, or MNase-seq), transcription factor (TF) binding, and/or histone marks known to be associated with transcriptional regulatory activity, such as H3K27ac, H3K4me1, and H3K4me3 (Maurano et al., 2012; Schaub et al., 2012). Intriguingly, daVs for a particular disease appear to be specifically enriched in CREs that appear active in disease-relevant cell types. For example, a study from Farh and colleagues (2015) examined the overlap of variants associated with 21 autoimmune diseases with six histone marks in multiple primary immune cell types and conditions (Farh et al., 2015). Importantly, the authors imputed the genotypes of variants not directly genotyped in their respective GWAS, and determined which variants were most likely to be causal using an algorithm that incorporates the LD structure and association pattern at each locus. The authors found that candidate causal variants were enriched in predicted B and T cell enhancers (consistent with the expected cellular origin of autoimmune diseases), and that this enrichment

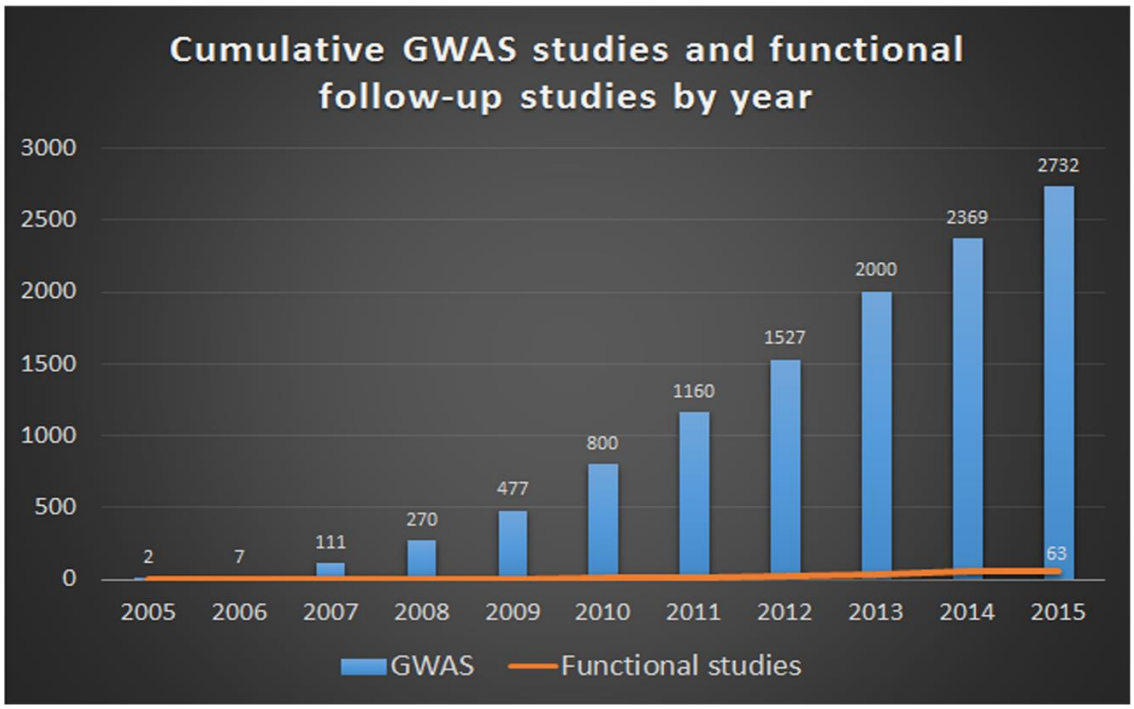


Figure 1.3. The mechanistic understanding of GWAS disease risk loci lags far behind the total number of published GWAS and SNP-trait associations.
The number of total GWAS published as of the indicated year (according to the EBI GWAS catalog (Welter et al., 2014)) are shown in blue, whereas the total number of published post-GWAS functional studies as of each year are shown by the orange line. 2015 data does not include studies published after 10/30/15. The number of published GWAS likely underestimates the number of unique SNP-trait associations, since many individual GWAS identify multiple disease or trait risk loci. Thus, a very small percentage of the thousands of SNP-trait associations have been investigated mechanistically. Post-GWAS functional studies were identified systematically by 1) reviewing the titles, and in some cases, abstracts, of all research articles published in 23* of the highest impact biomedical research journals between 10/3/13 and 10/30/15, and 2) PubMed searches using the keywords "causal variant" and "functional variant". Occasional additional studies were identified through references provided in review or primary research articles.
**American Journal of Human Genetics*, *Cancer Cell*, *Cell*, *Cell Reports*, *Cell Stem Cell*, *eLife*, *Genome Biology*, *Genome Research*, *Human Molecular Genetics*, *Molecular Cell*, *Nature*, *Nature Biotechnology*, *Nature Communications*, *Nature Genetics*, *Nature Medicine*, *Nature Neuroscience*, *Nature Structural & Molecular Biology*, *Neuron*, *PLOS Genetics*, *PNAS*, *Science*, *Science Translational Medicine*, *Stem Cell Reports*

increases with the likelihood that the variant is causal (Farh et al., 2015). When expanding this analysis to 18 additional traits and diseases and incorporating epigenetic data from additional cell and tissue types, the authors observed an enrichment of variants associated with neurological disease in predicted brain promoters and enhancers, whereas blood glucose risk variants were enriched in regulatory regions predicted to be active in pancreatic islets (Farh et al., 2015).

PMID	First Author	Locus	Trait	Identify all trait-associated variants	Variant prioritization methods	Test all putative causal variants	Experimental approaches
25865496	Lu,X	*ETS1*	Systemic lupus erythematosus	Y	Epigenomic, *in silico*, chromatin interactions	Y	EMSA, LC/MS, AS-ChIP-qPCR
25918370	He,H	*FOXE1/PTCSC2*	Papillary thyroid carcinoma	Y	Epigenomic, *in silico*	Y	AS-ChIP-qPCR, cellular reporter assays, 3C
26211971	Ogura,Y	*BNC2*	Adolescent idiopathic scoliosis	Y	Epigenomic	Y	Cellular reporter assays, EMSA
26211970	Vicente,CT	8q21	Asthma and hay fever	Y	Epigenomic	Y	AS-3C, cellular reporter assays, ChIP-qPCR
26287746	Claussnitzer,M	*FTO*	Obesity	PR	Epigenomic, *in silico*	N	Cellular reporter assays, genome editing, EMSA
26637976	Roman,TS	*GALNT2*	High density lipoprotein cholesterol levels	Y	Epigenomic	Y	Cellular reporter assays, EMSA, ChIP-qPCR
26398868	Spisak,S	6q22.1	Prostate cancer	Y	Epigenomic	Y	AS-ChIP-qPCR, epigenome editing, genome editing
26560027	Oldridge,DA	*LMO1*	Neuroblastoma	Y	N/A	Y	AS-ChIP-seq, cellular reporter assays
26695686	Wang,M	3q28	Bladder cancer	Y	*in silico*	N	Cellular reporter assays
26928228	Dunning,AM	*ESR1*	Breast cancer	Y	Epigenomic	Y	AS-3C, EMSA, cellular reporter assays, ChIP-qPCR
27096366	Soldner,F	*SNCA*	Parkinson's disease	PR	Epigenomic, *in silico*	N	Genome editing, ChIP-qPCR, EMSA
27055116	Nakaoka,H	9p21	Endometriosis	Y	Epigenomic	Y	AS-3C, AS-ChIP-qPCR
27162171	Wang,X		QT/QRS interval	PR	Epigenomic	Y	Cellular reporter assays, 4C
27149122	Smith,JG	5q22	Heart failure mortality	Y	Epigenomic	N	Cellular reporter assays
27148741	Vodo,D	*ST18*	Pemphigus vulgaris	Y	Epigenomic, *in silico*, conservation	N	Cellular reporter assays
27213290	Zheng,J	*LINC00673*	Pancreatic ductal adenocarcinoma	Y	*in silico*	N	Cellular reporter assays, isoform overexpression
27259051	Painter,JN	14q32	Endometrial cancer	Y	Epigenomic	Y	AS-3C, cellular reporter assays, EMSA, ChIP-qPCR
27402876	Wyszynski,A	2q35	Breast cancer	PR	Experimental (3C)	N	3C, AS-ChIP-qPCR, epigenome editing

27524613	Kandaswamy,R	15q15.1	Chronic lymphocytic leukemia	Y	Epigenomic, *in silico*, chromatin interactions, conservation	N	Cellular reporter assays, EMSA, 4C
27580880	Hoskins,JW	13q22.1	Pancreatic cancer	Y	Tested all variants	Y	3C, cellular reporter assays, EMSA, LC/MS
27539148	Xia,Q	*TCF7L2*	Type 2 diabetes	PR	N/A	N	Capture-C, 4C, genome editing
27601076	Lawrenson,K	19p13	Breast and ovarian cancer	Y	Epigenomic, experimental (3C)	Y	Cellular reporter assays, 3C, genome editing
27640304	Ghoussaini,M	5p12	Breast cancer	Y	Epigenomic	Y	3C, cellular reporter assays, EMSA
27655404	Powell,JE	1p36.12	Endometriosis	Y	Experimental (3C)	Y	3C, cellular reporter assays
27745831	Mika,KM	*TAP2*	Fecundability	N	N/A	N	Cellular reporter assays
27848966	Schmiedel,BJ	17q21	Asthma	PR	Epigenomic	N	ChIP-seq, ChIP-qPCR, 4C
27866707	Ye,J	4q25	Atrial fibrillation	N	Epigenomic, conservation	N	Transgenic reporter assays, cellular reporter assays, genome editing, EMSA, ChIP-qPCR
27817866	Vince,N	*MHC*	HLA-C levels	Y	*in silico*	N	EMSA, ChIP-qPCR

Table 1.1. Post-GWAS functional studies published in a recent 2 year period.

This list consists of studies published between March 1, 2015 and March 1, 2017, as identified by applying the methods described in **Figure 1.3** to the aforementioned 2 year time period. This approach identified 28 studies that investigated the mechanisms underlying a specific GWAS risk locus. In order to maximize the relevance of the table to this chapter, studies that investigated causal protein-coding variants were not included. The scientific rigor of each study is assessed in the last four columns, regarding 1) whether the study used statistical methods, such as imputation, to identify all trait-associated variants (Y=yes; N=no; PR=previously reported), 2) the methods by which the study prioritized candidate causal variants for experimental investigation, 3) whether the study experimentally tested all putative causal variants after prioritization for allele-specific function, and 4) the experimental methods used to test variants for allele-specific function. Prioritization approaches include epigenomic annotation of predicted transcriptional regulatory function, *in silico* prediction of transcription factor or microRNA binding sites, evolutionary conservation, and chromatin interactions between regions harboring candidate causal variants and known active regions, such as gene promoters. PMID=PubMed ID; AS=allele-specific, LC/MS=liquid chromatography/mass spectroscopy, which in all cases was used to identify peptides bound to a DNA probe in EMSA experiments.

Based on these results and other similar reports (Maurano et al., 2012), it has been suggested that many GWAS causal variants may influence disease risk by altering the function of cell type-specific regulatory elements, such as enhancers, with ensuing changes in target gene expression. This hypothesis is strongly supported by the overlap of daVs with expression quantitative trait loci (eQTLs) – specifically, variants that are associated with the expression (mRNA) levels of one or more genes are more likely to be daVs than would be expected by chance (Fu et al., 2012; Nicolae et al., 2010; Schaub et al., 2012). Furthermore, the cell type in which the eQTL effect is observed often matches cell types thought to be relevant to the disease in question, consistent with the overlap of daVs with disease-relevant tissue-specific CREs. In a study by Raj and colleagues (2014), a large-scale eQTL analysis in primary T cells and monocytes, representing adaptive and innate immune processes, respectively, was performed (Raj et al., 2014). The authors found a significant overlap between variants associated with expression in these cell types and variants associated with autoimmune diseases. Moreover, some daVs were only associated with gene expression levels in one of the two immune cell types. For example, daVs for Alzheimer's disease were associated with gene expression levels only in monocytes, which have been implicated in Alzheimer's pathogenesis and are related to the cellular precursors of brain microglia (Raj et al., 2014). While these studies support a role of cell type-specific *cis*-regulatory variation in complex disease pathogenesis, variants can also affect gene expression levels through post-transcriptional processes such as mRNA splicing and stability (Pai et al., 2015). Indeed, several studies have characterized functional daVs that may influence disease risk through these types of effects (Paraboschi et al., 2014; Richardson et al., 2013; Wang et al., 2014a; Zhang et al., 2011).

Other studies have associated genetic variants with altered levels of DNA methylation (mQTLs) (Hannon et al., 2016; Kaplow et al., 2015), DNase hypersensitivity (dsQTLs) (Degner et al., 2012), and TF binding (bQTLs) (Ding et al., 2014; Tehranchi et al., 2016), and some of these reports show significant overlap of these variants with daVs as well (Hannon et al., 2016; Tehranchi et al., 2016). Taken together, these observations suggest that many GWAS causal variants influence disease risk by altering the function of a CRE (presumably by affecting the recruitment of one or more *trans*-acting factor), which results in altered gene expression levels and disease-

relevant phenotypes. Notably, eQTL studies have consistently shown that most eQTL effects are of relatively small magnitude (<2-fold change in expression) (Dimas et al., 2009; GTEx Consortium, 2015), agreeing with the results of large-scale experimental characterizations of putative regulatory variants (Patwardhan et al., 2012; Tewhey et al., 2016). However, while much focus has been applied to identifying and characterizing functional *cis*-regulatory variants and their effects on gene expression, the mechanisms by which small changes in gene expression affect cellular or organismal phenotypes to influence disease risk are not well understood.

1.2.3 Recent advances in functional genomics

As mentioned previously, the number of SNP-trait associations established by GWAS has increased astronomically in the last decade. The number of SNP-trait associations that have been functionally dissected in an unbiased and comprehensive manner, however, is still relatively low (**Table 1.1**). Fortunately, this situation appears to be changing as a result of the advent of a series of experimental techniques and large-scale databases.

In the mid/late 2000s, there was limited information regarding the specific genetic variants that exist in human populations, and the functions of the noncoding regions of the genome were largely unexplored. Some impressive early post-GWAS studies relied upon deep sequencing or bacterial artificial chromosomes (BACs) to identify all the variants spanning a disease-associated haplotype, and nominated causal variants based on biochemical assays or cell culture-based experiments (Harismendy et al., 2011; Musunuru et al., 2010; Verlaan et al., 2009). For example, Musunuru et al. (2010) used BACs to identify all the variants located in a haplotype that had been associated with low density lipoprotein levels. The authors then used reporter assays to demonstrate haplotype-specific effects on gene expression, and, in combination with gel shift assays, identified the causal variant responsible for this effect (Musunuru et al., 2010). However, recent large-scale projects such as the International HapMap and 1000 Genomes projects (1000 Genomes Project Consortium et al., 2015; International HapMap 3 Consortium et al., 2010) have extensively characterized genetic variation in numerous human populations, obviating the need for sequencing (unless the causal variant is rare) and allowing for refinement and superior resolution

of association signals (Edwards et al., 2013). Many early post-GWAS studies also performed laborious epigenetic experiments in order to characterize potential CREs at their loci of interest (Harismendy et al., 2011; Zhou et al., 2012). However, as a result of large-scale data sets generated by The ENCODE Project (ENCODE Project Consortium et al., 2012), the NIH Roadmap Epigenome Project (Roadmap Epigenomics Consortium et al., 2015), the FANTOM consortium (Andersson et al., 2014; FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al., 2014), and others, there are now impressive annotations of putative CREs in hundreds of human cell types and tissues; therefore, such in-house experiments are oftentimes unnecessary. In addition, a wealth of eQTL data is now available for dozens of cell and tissue types (Albert and Kruglyak, 2015; GTEx Consortium, 2015), such that the association of a daV with gene expression levels can easily be queried, and potential causative genes can be identified. Finally, several new techniques have been developed that show great potential for the interrogation of allele-specific transcriptional processes, including genome and epigenome editing (Komor et al., 2016), chromosome conformation capture (3C)-based assays (Denker and de Laat, 2016), massively parallel reporter assays (MPRAs) (Inoue and Ahituv, 2015), and other allele-specific high-throughput sequencing (HT-seq) analyses.

1.2.4 A general framework for the functional dissection of a genetic risk locus

1.2.4.1 Statistical approaches

How might one leverage the wealth of genomic data that exists today in order to derive biological meaning from a GWAS-implicated disease risk locus? Here, I propose a general framework for translating a statistical association by GWAS to a functional understanding of the causal genetic variant(s), which is outlined in **Figure 1.4**. First, the resolution of microarray-based GWAS can be greatly increased by performing imputation of ungenotyped variants, using population-based sequencing data, such as that from the 1000 Genomes Project (Edwards et al., 2013). In this way, the significance of association of virtually all common (minor allele frequency $\geq$ 1%) variants with disease risk can be estimated (1000 Genomes Project Consortium et al., 2012). Conditional analyses can additionally be performed to determine if multiple weakly linked or

unlinked causal variants are contributing to the association of the same locus with disease risk (Edwards et al., 2013), as compared to a situation in which only one signal exists at the locus in question. In one example of the former situation, Glubb and colleagues (2015) performed a meta-analysis of breast cancer GWAS, finding a complex pattern of association involving at least three independent signals at and around the *MAP3K1* locus (Glubb et al., 2015). In an example of the latter situation, one study performed conditional analyses on seven loci associated with levels of adiponectin, an adipocyte-secreted protein associated with cardiovascular and metabolic traits (Wu et al., 2014). After conditioning on the sentinel GWAS SNP for each locus, six out of seven loci showed no residual association at any other variants, suggesting that these associations are driven by one or more strongly linked functional variants (Wu et al., 2014).

Next, most GWAS are performed initially in genetically similar groups of cases and controls, leading to the association of traits with haplotypes as defined in these genetic groups. In this context, trans-ethnic fine-mapping can be a useful approach to refine the region of association, thereby reducing the number of candidate causal variants, due to the reduced LD and smaller haplotype blocks in certain populations, particularly Africans (Campbell and Tishkoff, 2008; Edwards et al., 2013). For example, Guthridge et al. (2014) used such an approach, combined with re-sequencing of the candidate region, to reduce the number of candidate causal variants at a lupus-associated locus from 30 to 3 (Guthridge et al., 2014).

In addition, due to the massive amount of eQTL data now available (Albert and Kruglyak, 2015), daVs can easily be investigated for association with gene expression levels in many cell and tissue types. This is an important point, because if a daV is associated with the expression levels of a gene, a testable hypothesis regarding the function of the causal variant can be formulated. Importantly, conditional analyses can be performed using the sentinel GWAS and eQTL variants to determine if both effects are likely driven by the same underlying mechanism.

While studies integrating GWAS and eQTL data have reported that nearly half of all daVs are associated with gene mRNA levels in at least one cell type (Fu et al., 2012; Nicolae et al., 2010), there are several other mechanisms by which a functional variant could influence disease

risk. First, a variant could affect protein levels through effects on translation or protein stability without an effect on mRNA levels; indeed, up to 1/3 of variants that associate with protein levels (pQTLs) do not associate with the mRNA levels of the same gene (Battle et al., 2015), although few studies have examined this overlap (Hause et al., 2014; Wu et al., 2013). In addition, a GWAS causal variant may alter the amino acid sequence of a protein, thereby affecting protein function rather than abundance (Cooper and Shendure, 2011). These possibilities can usually be excluded, however, if there are no daVs in exonic regions. It is also increasingly being recognized that many daVs are located at long non-coding RNA (lncRNA) loci, implying that some variants influence disease risk through effects on lncRNA levels or function. The annotation and functional characterization of lncRNAs is a rapidly growing field, and there are already examples of functional daVs that affect lncRNA function (Jendrzejewski et al., 2012; Zheng et al., 2016).

1.2.4.2 Incorporating public functional genomics data

Notwithstanding the caveats mentioned above, if a sentinel GWAS SNP and any strongly linked variants (e.g. $r^2 \geq 0.8$) are 1) associated with the mRNA levels of one or more genes in a disease-relevant cell type, and 2) are all intronic or intergenic, then it is reasonably likely that the causal variant influences disease risk by modulating the transcriptional regulation of the target gene(s). Given the massive amount of publicly available epigenomic data mentioned previously, researchers are increasingly using these data sets to prioritize candidate causal variants based on overlap with accessible chromatin, TF binding, and/or histone marks associated with regulatory activity (**Figure 1.4**). In addition, the pattern of histone modifications observed at a putative CRE can help predict which type of regulatory element it may be (e.g. promoter, enhancer, insulator, etc.), and therefore, how the region should be functionally tested. One study used such an approach to identify functional variants underlying the association of the 8q21 locus with allergic diseases (Vicente et al., 2015). The sentinel GWAS SNP was found to be associated with the expression of the *PAG1* gene in B lymphoblasts, and the authors used ENCODE data to identify 35 SNPs (out of a total of 118 that are in moderate LD ($r^2 \geq 0.6$) with the sentinel SNP) that overlap

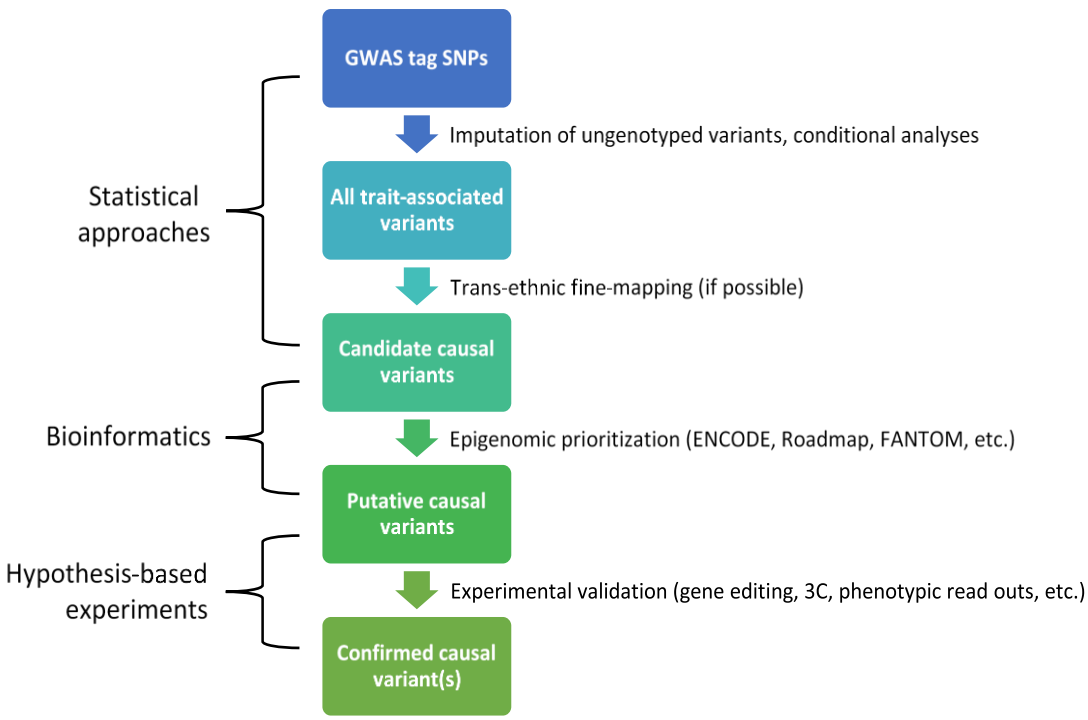


Figure 1.4. A framework for the functional dissection a GWAS-nominated disease risk locus.
SNPs associated with disease risk by GWAS (tag SNPs) can be expanded upon with imputation, and independent disease risk associations can be identified with conditional analyses. Candidate causal variants can be prioritized through trans-ethnic fine-mapping, and, if the variants are associated with gene expression levels, epigenomic prioritization. Various experimental approaches can then be used to confirm which variants affect regulatory function. While this pipeline will not identify exonic causal variants that might affect mRNA stability or protein function, such variants likely do not underlie the majority of GWAS risk loci.

four distinct regions of DNase I hypersensitivity and enhancer-associated histone marks in this cell type. These potential regulatory regions were further investigated using multiple approaches, including chromosome conformation capture (3C) and reporter gene assays (Vicente et al., 2015).

In addition, epigenomic data sets have been used to investigate loci that are associated with disease risk by GWAS, but do not reach statistical significance after correction for multiple hypothesis testing (Wang et al., 2016). Indeed, GWAS typically employ the Bonferroni correction method for multiple hypothesis testing, which is likely to be overly conservative due to LD between nearby SNPs throughout the genome (Gao et al., 2010a; Johnson et al., 2010). Thus, some SNPs

that do not reach the accepted genome-wide significance threshold (typically $P<5x10^{-8}$) may represent true disease risk loci. To investigate this hypothesis, a study by Wang et al. (2016) examined the overlap of SNPs associated with cardiac QT interval with epigenetic enhancer marks in cardiac and non-cardiac tissues. The authors found that both genome-wide significant SNPs ($P<5x10^{-8}$) and "sub-threshold" SNPs ($5x10^{-8}\leq P\leq 1x10^{-4}$) were significantly enriched in predicted cardiac enhancers, and >70% of enhancers harboring sub-threshold SNPs exhibit allele-specific regulatory activity in induced pluripotent stem cell (iPSC)-derived cardiomyocyte luciferase reporter assays (Wang et al., 2016). Furthermore, enhancer-associated sub-threshold SNPs were more strongly associated with QT interval than non-enhancer-associated sub-threshold SNPs, and the enhancer-associated SNPs were more likely to reach genome-wide significance in larger GWAS meta-analyses (Wang et al., 2016). Thus, prioritizing candidate causal variants based on epigenomic annotations is a reasonable method for functionally characterizing candidate disease risk loci.

Not long after these large-scale epigenomic data sets became available, many groups developed online resources that combined these data with other useful data sets, including those from eQTL studies, enhancer target gene predictions, phylogenetic conservation scores, and *in silico* TF motif prediction tools. There are now a multitude of available web-based tools that are useful for prioritizing candidate functional *cis*-regulatory variants at loci of interest, which have been recently reviewed elsewhere (Tak and Farnham, 2015). Typically, a variant or list of variants can be queried for overlap with potential CREs, disruption of TF binding motifs, association with gene expression levels, and other features that can help predict the likelihood of that given variant being functional.

1.2.4.3 Testing the function of a regulatory variant

Once a list of candidate CREs is identified, all containing one or more potential causal variants, various experimental approaches can be used to test the functions of these regions. A common approach involves *in silico* analysis to determine if a particular variant is predicted to disrupt a TF binding motif (Heinemeyer et al., 1998; Mathelier et al., 2014), but evidence suggests

that many causal variants that may in fact disrupt TF binding may not reside in known TF motifs. For example, the Farh et al. study estimated that only 10-20% of autoimmune GWAS causal *cis*-regulatory variants reside in known TF motifs (Farh et al., 2015). An alternative approach is to functionally test all candidate CREs, using both the risk and protective alleles of the candidate causal variants. Cell culture-based reporter assays have been widely used for these purposes: the candidate CRE is cloned into a physiologically relevant position with respect to the reporter gene, and transfected into a relevant cell type, and the activity of CREs containing alternate alleles (or haplotypes) are compared. Since some CREs are not only cell type-specific, but signal-dependent (Dickel et al., 2014; Shlyueva et al., 2014), it is necessary to determine the appropriate experimental conditions in which to test the variant in order to avoid false positives and false negatives.

Rather than testing reporter constructs one-by-one in cell culture contexts, several groups have developed massively parallel reporter assays (MPRAs), in which thousands of variants can be tested in a single experiment (Inoue and Ahituv, 2015; Melnikov et al., 2012; Patwardhan et al., 2012; Shen et al., 2016; Smith et al., 2013; Ulirsch et al., 2016; Vockley et al., 2015). These approaches will surely be useful for researchers investigating large numbers of loci and variants at once. For example, Tewhey and colleagues (2016) tested ~30,000 SNPs representing >3,500 eQTL signals (eSNPs), testing each eSNP and all variants in perfect LD with it for enhancer activity in immortalized liver (HepG2) and B lymphoblast cell lines. Roughly 12% of the putative CREs displayed enhancer function in one or both of the cell types tested, and of these, ~25% contained SNPs that caused significant changes in reporter gene expression (Tewhey et al., 2016). Importantly, ~80% of the expression differences caused by these variants agreed with the direction of previously published eQTL effects in the same cell type (Tewhey et al., 2016). In addition, the majority of functional variants identified in this study altered reporter gene levels by less than 2-fold, consistent with eQTL effect sizes predicted by previous studies (Dimas et al., 2009; GTEx Consortium, 2015; Patwardhan et al., 2012). These results reinforce current thinking that most

causal gene regulatory variants underlying complex traits and diseases have modest effects on target gene expression.

While reporter assays are often a useful first step in determining the function of a potential regulatory variant, they have several limitations: first, reporter assays can display a significant amount of transcriptional noise, and thus are not always reproducible (Brown et al., 2013). Second, small differences in reporter activity can result from small differences in the molar amounts of each plasmid that is transfected into cells, which is unavoidable even with the most accurate DNA concentration measurements. These issues can make small differences in expression difficult to distinguish statistically. Perhaps most importantly, reporter assays test the transcriptional function of a variant in the context of plasmid DNA, rather than the native genomic context in which the variant actually exists (Inoue and Ahituv, 2015). This situation can produce false negative and false positive results, due to the intricate relationships between DNA, histones, transcription factors, and long-range chromatin interactions (Inoue and Ahituv, 2015).

In light of these issues and others, a more physiologically-relevant method to confirm the function of a regulatory variant may be genome editing (Engel et al., 2016). There are currently three available genome-editing technologies: the protein-based zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs), and the nucleic acid-based clustered regularly interspaced short palindromic repeat (CRISPR)-based systems (Gaj et al., 2016). Due to its simple, cheap, and rapidly scalable utility, CRISPR has become the most widely used genome editing method. With genome editing, a researcher can theoretically make any change he/she wants to any region of the genome, and then use well-established molecular assays to determine if changes in regulatory function and gene expression have occurred. In one of the first applications of genome editing to a GWAS-nominated locus, Bauer and colleagues (2013) used TALENs to delete a 10kb intronic region in the mouse ortholog of *BCL11A*. The orthologous region in humans harbors the top SNPs associated with fetal hemoglobin levels, for which *BCL11A* is a known repressor (Bauer et al., 2013). Thus, the causal variant may function by regulating *BCL11A*, which might in turn affect levels of fetal and embryonic β-globin – and indeed, manipulating this pathway

is an attractive prospect for treating β-hemoglobinopathies (Lettre and Bauer, 2016). The authors demonstrate that some of these top SNPs fall within three distinct regions of open chromatin and enhancer-associated histone marks that are specific to human erythroid cells, consistent with the erythroid-specific expression patterns of the globin genes. The SNP most significantly associated with fetal hemoglobin, located in one of these three candidate CREs, disrupts a DNA motif known to bind the erythroid TFs GATA1 and TAL1, and the authors demonstrate that the motif-disrupting allele indeed reduces binding of these factors in multiple primary human erythroid samples. (Bauer et al., 2013). Next, the authors used TALENs to delete the 10kb interval containing the three candidate CREs (including the putative causal variant) in mature murine erythroid cells. This resulted in drastically reduced Bcl11a expression levels and drastically increased embryonic β-globin levels, thus establishing the region as a functional *Bcl11a* enhancer that is required for repression of embryonic β-globin (Bauer et al., 2013).

In the example above, a large genomic region was deleted to demonstrate the importance of that region to gene regulatory function. However, genome editing can also be used to make more precise changes, such as mutating an individual SNP from one allele to the other. In Spisák et al. (2015), the authors used TALEN-mediated homology directed repair (HDR) to confirm the functional role of a SNP previously reported to influence prostate cancer risk by modulating the expression of the *RFX6* gene (Huang et al., 2014). Specifically, they compared edited and unedited prostate cancer cell line clones, and demonstrated that this candidate causal variant altered *RFX6* expression levels by ~2-fold (Spisak et al., 2015). Moreover, the authors characterized the regulatory potential of the region harboring the SNP by fusing a catalytically-inactive TALE array with either a VP64 transcriptional activation domain, or LSD1, a histone lysine-specific demethylase known to remove H3K4 methylation enhancer marks and decrease enhancer activity (Mendenhall et al., 2013). As expected, site-specific recruitment of VP64 and LSD1 to the putative causal SNP increased and decreased *RFX6* levels, respectively, establishing the region harboring the causal variant as a bona fide regulatory element (Spisak et al., 2015). Thus, genome editing technologies

can also be used to validate potential CREs by altering their epigenetic state (referred to as "epigenome editing") (Dominguez et al., 2016).

A more recent study used CRISPR/Cas9 gene editing to investigate candidate causal variants at the *SNCA* locus (Soldner et al., 2016), which is associated with risk for Parkinson's disease (PD) (Nalls et al., 2014), and which encodes α-synuclein, the protein that accumulates in the characteristic Lewy Body inclusions of PD (Toulorge et al., 2016). The authors demonstrate that the *SNCA* risk haplotype is associated with increased *SNCA* brain expression, which has previously been associated with PD pathogenesis, since families with duplications or triplications of the *SNCA* locus (resulting in increased *SNCA* levels) exhibit Mendelian forms of PD (Singleton et al., 2003). After prioritizing candidate causal variants based on epigenetic signatures and *in silico* TF motif predictions, the authors deleted a 500bp putative enhancer at this locus containing two SNPs in human embryonic stem (ES) cells (Soldner et al., 2016). They reinserted the 500bp region using HDR with either the risk or protective alleles of the two SNPs, and differentiated the ES cells into neural precursors and mixed neuronal cultures (Soldner et al., 2016). Cell clones bearing the risk-associated alleles of the enhancer SNPs demonstrated significantly higher *SNCA* levels than clones bearing the protective alleles, and this effect was driven entirely by the variant predicted to be functional by *in silico* and experimental analyses (Soldner et al., 2016).

Analogous to the high-throughput reporter assays (MPRAs) mentioned above, several groups have also developed high-throughput CRISPR screens to identify essential genes (Wang et al., 2014b), potential drug targets (Shi et al., 2015) or noncoding regulatory regions (Diao et al., 2016; Fulco et al., 2016; Korkmaz et al., 2016; Rajagopal et al., 2016; Sanjana et al., 2016). As with MPRAs, high-throughput gene editing screens will facilitate efficient testing of many candidate *cis*-regulatory regions and their associated variants in a single experiment, with the additional benefit of testing variants in their endogenous genomic context.

1.2.4.4 Determining the target gene(s) of a regulatory variant

While genome editing can confirm the allele-specific functions of a distal CRE, editing experiments alone cannot determine the mechanisms by which these elements affect transcription

of their target genes. The last few years have seen an explosion in the number of studies investigating how the genome is organized in the nucleus, both at small and large scales, and there is now abundant evidence that chromosomes can bend and form loops at kilobase and megabase scales, and that these loops play an important role in transcriptional regulation and disease (Merkenschlager and Nora, 2016; Pombo and Dillon, 2015; Wang and Dostie, 2016). While the transcription of a gene occurs at the promoter, enhancers and other distal regulatory elements appear to affect transcription by physically interacting with their target promoters, and oftentimes with each other, through chromatin looping (Merkenschlager and Nora, 2016; Pombo and Dillon, 2015; Wang and Dostie, 2016). Thus, physical contact between a distal regulatory element and a promoter may be considered evidence for a regulatory function of that element. The marriage of chromosome conformation capture (3C) with high-throughput sequencing has allowed for the investigation of all long-range contacts in the genome (Hi-C, an "all-versus-all" approach), or, with superior depth and resolution, all long-range contacts involving a region of interest, such as a gene promoter (4C, Capture-C, or Capture Hi-C; "one-versus-all" approaches) (Denker and de Laat, 2016).

The value that these approaches possess for post-GWAS functional studies (and for the study of eukaryotic transcriptional regulation in general) cannot be overstated. Even in cases in which the likely causal variant is already known based on statistical association, allele-specific effects on reporter genes and TF binding, etc., it can be extremely difficult to know *a priori* what the target gene(s) of the CRE harboring the variant might be (Edwards et al., 2013). "One-versus-all" approaches – in which a specific genomic region is captured or selectively amplified in conjunction with all interacting regions – are well-suited to identify the target promoters of enhancers and other distal elements, or conversely, all distal regulatory elements that interact with a given promoter, such that the regulatory effects of a variant can be linked to the correct gene(s).

A striking example of the importance of considering the three-dimensional organization of chromatin concerns the association of intronic genetic variants at the *FTO* locus with obesity (Smemo et al., 2014). The *FTO* locus is the strongest known risk factor for obesity, with an odds

ratio of >1.4 for the sentinel SNP located on chromosome 16q12.2 (Berndt et al., 2013). Furthermore, FTO expression levels have been reported to affect body mass and composition in mice; thus, *FTO* was considered by many to be the gene responsible for conferring risk for obesity at this locus (Church et al., 2010; Fischer et al., 2009; Gao et al., 2010b). However, while the top obesity-associated SNPs at this locus are all intronic, suggesting a regulatory effect for the causal variant, none of the SNPs have been associated with *FTO* expression (Smemo et al., 2014). To resolve this conundrum, Smemo and colleagues (2014) performed 4C in mouse embryos and brain to determine if the obesity-associated interval interacts with any genes other than *FTO*. This analysis demonstrated strong interactions between the obesity-associated region and the *Irx3* promoter, located several hundred kilobases away. As the obesity-associated region displays enhancer-associated histone marks, the authors then demonstrated enhancer activity for this region using transgenic mouse assays (Smemo et al., 2014). Importantly, they also demonstrate an association between the obesity-associated SNPs and *IRX3* expression in a large set of human brain samples, confirming *IRX3* as a likely target gene of the *FTO* enhancer region. These results were then corroborated by mouse models demonstrating a role for *Irx3* in body weight maintenance (Smemo et al., 2014). An impressive follow-up investigation by Claussnitzer et al. (2015) identified the likely causal variant at this locus, using precise genome editing and several other approaches (Claussnitzer et al., 2015). This work identified an additional target gene of the obesity-associated region, *IRX5*, which also appears to affect obesity-related cellular phenotypes. Thus, 3C-based approaches were essential in identifying the genes responsible for conferring obesity risk through long-range regulatory effects of obesity-associated variants at the *FTO* locus.

It has been suggested that if the activity of a distal regulatory region is modified by a functional variant, and regulation of the target gene(s) by such a region involves long-range interactions, then regulatory variants might influence the long-range interactions themselves (Dixon et al., 2015). Accordingly, 3C-based approaches have been used to identify allele-specific long-range interactions, typically using cell lines or tissues that are heterozygous for the disease-associated haplotype. It is worth emphasizing that allele-specific effects in general may be best

investigated within multiple heterozygous samples, rather than between samples of different genotypes, due to significant inter-individual phenotypic variation. In line with this, even robust allele-specific effects in heterozygous cell lines have been reported to be highly variable, and not well reproducible, across individuals (GTEx Consortium, 2015).

Allele-specific long-range interactions can be detected by using SNP-specific primers or probes for PCR-based approaches (e.g., 3C or 4C) (Holwerda et al., 2013), or, for approaches involving HT-seq (e.g. 4C or Hi-C), designing the experiment such that haplotype marker SNPs are present in the sequenced ligation products (Dixon et al., 2015; Stadhouders et al., 2014; Tang et al., 2015). In an example of the former approach, Stadhouders et al. (2014) used K562 cells to investigate long-range interactions involving a putative enhancer region that harbors variants associated with fetal hemoglobin levels (Stadhouders et al., 2014). The authors performed 3C with a SNP-specific primer in order to demonstrate allele-specific chromatin looping between the putative enhancer and the promoter of the *MYB* gene, which is a key hematopoietic and erythropoietic TF (Stadhouders et al., 2014). With regards to the latter approach, several groups have attempted to identify allele-specific interactions by sequencing the 3C ligation products and determining whether SNPs contained in the ligated fragments deviate from either a 50/50 allelic ratio, or the allelic ratio present in a control sample or condition (Dunning et al., 2016; Ghoussaini et al., 2014; Vicente et al., 2015). However, such marker SNPs are oftentimes not available, and, depending on the approach used and the experimental design, well-placed SNPs may lack sufficient sequencing depth to distinguish modest allele-specific interactions (Dixon et al., 2015). Furthermore, the relationship between allele-specific expression and allele-specific long-range interactions is not well understood, so further studies are needed to investigate these phenomena. Notably, while daV-containing CREs have been reported to engage in allele-specific long-range interactions, such as enhancer/promoter interactions (Dunning et al., 2016; Nakaoka et al., 2016; Painter et al., 2016; Stadhouders et al., 2014; Vicente et al., 2015), the potential involvement of causal GWAS variants in altering higher-order chromatin architecture is unknown.

One important caveat with respect to 3C-based approaches is that although distal CREs appear to engage in long-range interactions to influence transcription, these interactions may not always be necessary, nor sufficient, for gene regulation (Ghavi-Helm et al., 2014). Thus, 3C-based approaches should complement other experimental approaches to demonstrate the functionality of a candidate causal variant. In summary, 3C-based approaches are increasingly being used to complement post-GWAS functional studies, and the combination of genome editing with 3C and high-throughput sequencing is a powerful tool for dissecting the functions of *cis*-regulatory variants.

1.2.4.5 Determining the molecular function of a regulatory variant

If one or more of the approaches mentioned above succeed in identifying a functional *cis*-regulatory variant, the question remains as to how the variant affects the function of the CRE at the molecular level. Given the overwhelming evidence supporting the critical role of TFs and chromatin remodelers in transcriptional regulation (Lelli et al., 2012), coupled with the significant overlap of daVs with mQTLs and bQTLs (Hannon et al., 2016; Tehranchi et al., 2016), one may hypothesize that a causal *cis*-regulatory variant enhances or represses the ability of one or more trans-acting factors to bind the CRE. This effect may be direct (e.g., directly affecting binding of one or more TFs or chromatin-modifying enzymes) or indirect (e.g., affecting DNA methylation). Some variants may be located in confirmed TF binding sites from resources such as ENCODE, which in combination with motif analysis, can help predict which TFs may be affected by the variant. The effect of a variant on TF binding can be confirmed by chromatin immunoprecipitation quantitative polymerase chain reaction (ChIP-qPCR) using allele-specific probes or primers, such that allelic differences in TF binding can be determined at the variant of interest in a heterozygous cell line or tissue. This approach was used to confirm allele-specific binding of the transcription factor HOXB13 at a putative causal variant at the *RFX6* locus (Huang et al., 2014). Alternatively, ChIP-seq experiments can be performed to investigate potential allele-specific TF binding. In samples heterozygous for the candidate functional variant, sequencing reads covering the variant (or a nearby linked variant) are expected to be present in equal allelic ratios if the variant does not affect binding of that particular factor; conversely, deviations from a 50/50 allelic ratio have been used to

infer function of candidate causal variants both at pre-determined loci (Bauer et al., 2013) and in genome-wide analyses (Maurano et al., 2015). These ratios are often normalized to the allelic ratios present in the input DNA, in order to correctly interpret results.

Electrophoretic mobility shift assays (EMSAs, or "gel shift assays") are another method of determining whether a variant affects recruitment of a nuclear factor *in vitro*, although these assays can be difficult to interpret and lack biological context (Hellman and Fried, 2007). Antibodies raised against candidate TFs can be added to the reaction to confirm their binding to the variant; alternatively, purified recombinant protein can be used instead of nuclear extract to assess binding of specific TFs (Hellman and Fried, 2007). In many cases it may be difficult to predict which TFs or other types of nuclear proteins are affected by the variant, in which case unbiased approaches such as EMSA followed by mass spectrometry can be helpful (Stead et al., 2006). For example, a study by Fogarty and colleagues (2014) combined these approaches to investigate a putative causal variant at the *CDC123/CAMK1D* Type 2 diabetes risk locus (Fogarty et al., 2014). After prioritizing candidate causal variants based on epigenomic annotations and identifying a variant that affects enhancer activity in cell-based reporter assays, the authors performed EMSAs to determine which trans-acting factors might be affected by the variant. Twenty-one base pair probes containing either the risk or protective allele of rs11257655, the candidate causal variant, were incubated with nuclear extract from HepG2 immortalized liver cells and two rodent insulinoma cell lines. In all three extract types, the authors observed one or more risk allele-specific probe/protein complexes that could be supershifted with antibodies raised against the enhancer-binding proteins FOXA1 and FOXA2 (Fogarty et al., 2014). Consistently, rs11257655 is located within a predicted FOXA1/2 motif, and the protective allele alters a highly conserved "T" base pair within the motif. The authors further confirmed risk allele-specific binding of FOXA2 by performing a DNA affinity capture assay followed by mass spectroscopy (Fogarty et al., 2014).

1.2.4.6 Linking gene expression changes to complex traits

While determining the molecular mechanism by which a disease-associated variant affects gene expression is an intriguing line of inquiry, perhaps a more physiologically- and disease-

relevant question is to ask how small changes in a gene's expression levels affect cellular and organismal phenotypes in order to influence disease risk. Indeed, while many studies have reported genetic variants that alter *cis*-regulatory function, the mechanisms by which the resulting alterations in gene expression influence disease risk are often not investigated, or are unknown (Deplacnke et al., 2016). Some studies have functionally linked expression levels of the causative gene to disease-relevant phenotypes, but many of these studies relied upon imprecisely-controlled overexpression, strong knockdown, or knockout approaches (Huang et al., 2014; Kapoor et al., 2014; Musunuru et al., 2010; Smemo et al., 2014). Recapitulating the gene expression differences relevant to a disease risk locus is difficult for at least two reasons: first, eQTL effect sizes are, in terms of fold expression change, typically unknown, not reported, or small; second, it is technically difficult to finely titrate the overexpression or knockdown of a gene. To overcome this issue, some studies have looked for correlations between the expression levels of the gene of interest and disease-relevant phenotypes, across samples or individuals. For example, Huang et al. (2014) characterized a functional variant at the *RFX6* locus, which appears to increase prostate cancer risk by increasing enhancer-mediated *RFX6* regulation. Knockdown of *RFX6* impaired prostate cancer cell migration and invasion, and consistently, *RFX6* expression levels were positively correlated with tumor aggressiveness and relapse across 128 prostate cancer samples (Huang et al., 2014). Some studies have also reported trait-relevant phenotypes that are distinguishable between cell lines of different genotypes, such as pigmentation in melanocytes (Visser et al., 2012), although this may not be a common phenomenon among complex traits.

The limitation of the approaches discussed above is that they are correlational. Thus, to determine the phenotypic effects of allele-specific changes in gene expression, genome editing may again be the best approach. By mutating the causal variant from one allele to the other, the resulting changes in gene expression are 1) more likely to be physiologically relevant than those seen in overexpression or knockdown experiments, and 2) can be causally linked to the variant. By using HDR to mutate the *RFX6* causal variant, Spisák et al. (2015) demonstrate a 2-fold expression difference between risk and protective allele homozygote clones (Spisak et al., 2015).

Intriguingly, protective allele homozygote clones displayed notably different cellular morphology and impaired cellular adhesion compared with risk allele homozygotes. However, no effects on cell migration or invasion were seen, presumably because of the smaller changes in *RFX6* expression than in the Huang et al. study (Spisak et al., 2015). In Claussnitzer et al. (2015), precise editing of the *FTO* obesity causal variant in adipocytes not only caused the expected changes in target gene expression, but also affected metabolic rate, oxygen consumption, and thermogenesis, all pathways that are associated with obesity (Claussnitzer et al., 2015). While these initial results are promising, a more complete understanding of the mechanistic link between allele-specific changes in gene expression and risk for complex diseases and traits is needed.

1.2.5 Summary

As mentioned previously, GWAS have identified thousands of SNP-trait associations throughout the genome, linking common genetic variation to hundreds of complex diseases and traits. However, only a small fraction of these statistical associations have been thoroughly investigated to determine 1) which variant or variants are causal, 2) what the molecular functions of the causal variants are, 3) which genes are affected by the causal variants, and 4) how changes in the function or regulation of the causal genes lead to altered disease risk. This is also the case, specifically, for neurodegenerative disease. GWAS have identified ~200 loci associated with the four major neurodegenerative diseases (AD, PD, ALS, and FTLD) and related phenotypes (Welter et al., 2014), yet only one of these loci, the *SNCA* locus, which has been known to be involved in PD risk for decades (Polymeropoulos et al., 1997; Spillantini et al., 1997), has been mechanistically investigated in detail. Thus, to date, GWAS approaches have not significantly improved our understanding of the pathogenesis of neurodegenerative disease.

Given the emergence of the new tools, technologies, and experimental techniques mentioned in this section, researchers studying neurodegenerative disease are in a better position than ever to translate statistical associations by GWAS into a functional understanding of disease mechanisms. In the case of FTLD, a fatal and untreatable neurodegenerative disease that affects patients as early as the 5th decade of life, common variants at the *TMEM106B* locus have been

associated with risk for FTLD-TDP, hippocampal sclerosis, TDP-43 pathology in Alzheimer's disease, and other neurological phenotypes (Nelson et al., 2015; Rutherford et al., 2012; Vass et al., 2011; Yu et al., 2015). In addition, *TMEM106B* genotype affects age at death and/or age at onset of disease in FTLD-TDP cases caused by mutations in the *GRN* and *C9orf72* genes (Cruchaga et al., 2011; Finch et al., 2011; Gallagher et al., 2014). Thus, *TMEM106B* is not only a risk factor for FTLD, but a genetic modifier that can affect disease manifestation. Understanding how genetic variation at *TMEM106B* affects risk and prognosis of FTLD-TDP could provide critical insights into the pathogenesis of FTLD, and possibly other neurodegenerative diseases.

1.3 PART III: *TMEM106B* is a genetic risk factor for frontotemporal lobar degeneration

As mentioned in Chapter 1.2, many human diseases, including diseases of aging, are complex traits. FTLD is no exception, with up to ~90% of cases not attributable to Mendelian mutations (Benussi et al., 2015). Like any complex trait, the majority of FTLD cases are likely caused by a combination of genetic and environmental risk factors. However, compared to other neurodegenerative diseases, FTLD has a relatively high heritability (30-50% of patients have a family history of dementia (Bang et al., 2015; Tan et al., 2017)), making it an attractive disease to study various types of genetic effects. In addition, FTLD is currently untreatable and fatal, so the identification of pathways involved in the pathogenesis of the disease is urgently needed.

In this section, I will review the discovery and subsequent investigations of a locus on chromosome 7p21 first associated with FTLD by GWAS, thus setting the stage for the work described in this dissertation.

1.3.1 Discovery of *TMEM106B* as a risk locus for neurodegeneration

1.3.1.1 Association of the *TMEM106B* locus with frontotemporal lobar degeneration

The first FTLD GWAS included 515 FTLD-TDP cases of Western European ancestry, 499 of which had autopsy-confirmed TDP-43 pathology, and 16 of which were living patients with *GRN* mutations, who are known to develop TDP-43 pathology (Van Deerlin et al., 2010). The authors hypothesized that increasing the neuropathological homogeneity of cases (all cases exhibited TDP-43 pathology) would compensate for the small sample size (GWAS typically consist of many thousands of cases and controls). Of the 515 cases, a total of 89 had *GRN* mutations (16 living *GRN*+ FTLD-TDP individuals and 73 with neuropathologically-confirmed TDP-43 pathology), 80 had *C9orf72* mutations (this was unknown at the time of the GWAS since *C9orf72* was not identified as an FTLD-causing gene until 2011), and the remaining cases tested negative for *MAPT* and *TARDBP* mutations, the former of which causes non-TDP-43 pathology. The cases were amassed from 45 institutions in 11 countries (United States, Canada, United Kingdom, Netherlands, Belgium, Spain, Germany, Australia, Finland, France and Sweden). The ethnically-matched control cohort

consisted of 1,297 children from the Children's Hospital of Philadelphia, and 1,212 adults from the Wellcome Trust Case Control Consortium. The inclusion of children and the lack of neurological confirmation of normal brain function in the adults was justified by the fact that FTLD is relatively rare (~10-20 per 100,000 individuals in the 45-64 age group), such that the genotype frequencies in any ethnically-matched control group would very likely be representative of the population in general.

Out of a total of >500,000 genotyped SNPs, the GWAS identified three SNPs on chromosome 7p21 that reached a genome-wide statistically significant ($P<5x10^{-8}$) association with FTLD-TDP (**Figure 1.5**). The most significant (sentinel) SNP was rs1990622 ($P=1.08x10^{-11}$), which is located ~7kb downstream of the gene *TMEM106B*. A second SNP, rs6966915 ($P=1.63x10^{-11}$), is in complete linkage disequilibrium (LD) ($r^2=1$) with rs1990622 and is located in the 5th intron of *TMEM106B*. The third SNP, rs1020004 ($P=5.00x10^{-11}$), is in moderate LD ($r^2=0.7$) with both rs1990622 and rs6966915 and is located in the 3rd intron of *TMEM106B* (**Figure 1.5**). The major alleles of all three SNPs were associated with increased risk of FTLD-TDP, with an odds ratio of 1.64 for the risk allele of rs1990622, and have ~60% allele frequencies in populations of Western European ancestry (1000 Genomes Project Consortium et al., 2015). Interestingly, the association was stronger in individuals with *GRN* mutations (*GRN*+ FTLD-TDP) ($P=1.34x10^{-9}$ for rs1990622) than in individuals without *GRN* mutations (*GRN*- FTLD-TDP) ($P=6.90x10^{-7}$ for rs1990622), despite the much smaller number of *GRN*+ individuals (n=80 *GRN*+ FTLD-TDP versus 435 *GRN*- FTLD-TDP). Thus, genetic variation at the *TMEM106B* locus may affect FTLD pathogenesis even in the presence of a highly penetrant mutation (Van Deerlin et al., 2010).

The association of rs1990622 and rs1020004 with FTLD-TDP was replicated in an independent cohort of 89 FTLD-TDP cases and 553 self-identified European-descended controls ($P=0.002$ for rs1990622 and $P=0.0004$ for rs1020004; rs6966915 genotyping was not performed due to technical issues) in the original GWAS study (Van Deerlin et al., 2010). The replication cohorts consisted of cases and controls not included in the GWAS, but that came from the same sources and met the same criteria as the GWAS cohorts. In addition, when using the combined

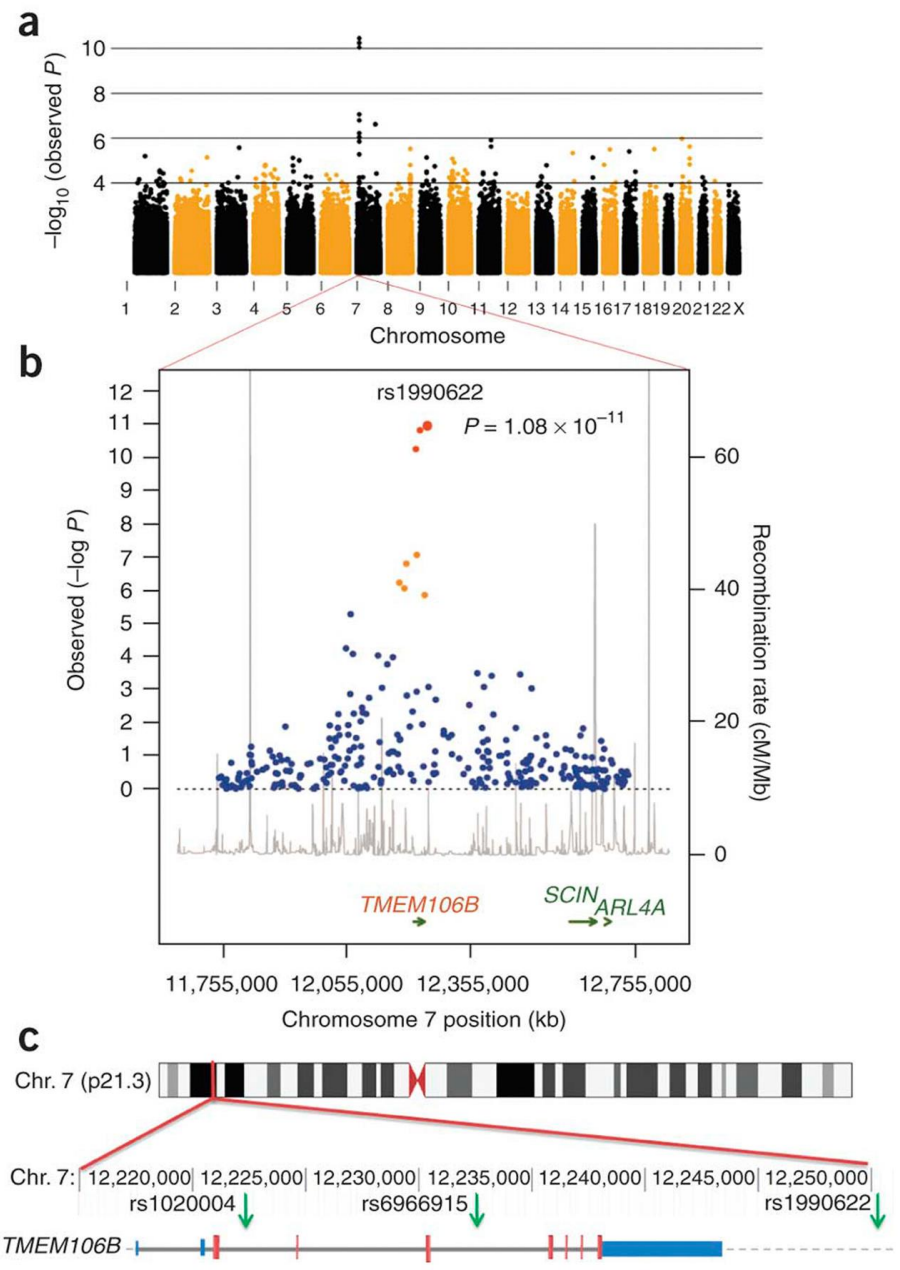


Figure 1.5. Association of common variants at the 7p21 locus with FTLD (from Van Deerlin et al., 2010, *Nature Genetics*).
(A) Manhattan plot of $-\log_{10}$(observed *P*-value) for all genotyped SNPs across genome demonstrating region of genome-wide significant association on chromosome 7.
(B) Regional plot of the *TMEM106B* locus. **Foreground plot:** Scatter plot of the SNP $-\log_{10}$ *P*-values plotted against physical position (NCBI build 36). **Background Plot:** Estimated recombination rates (from phase 2 of the HapMap) plotted to reflect the local LD structure. The color of the dots represent the strength of LD between the top SNP rs1990622, and its proxies (red: $r^2 \geq 0.8$; orange $0.8 > r^2 \geq 0.4$; blue <0.4). Gene annotations were obtained from assembly 18 of the UCSC genome browser.
(C) Location of the 3 most strongly associated SNPs (green arrows) relative to the gene structure of *TMEM106B* (blue bars, 3′ and 5′-untranslated regions; larger red bars, coding exons; thick gray line, intronic regions; gray dashed line, downstream chromosome sequence) and chromosome 7 location.

GWAS cohort, the risk allele of rs1020004 was associated with shorter duration of (i.e., more severe) disease. Importantly, an additional replication using 192 cases of non-specific FTLD (neuropathology not known) found no association between any of the three SNPs and FTLD, highlighting the importance of the authors' initial decision to control for neuropathological subtype (Van Deerlin et al., 2010).

Three studies in 2011 attempted to independently replicate the 7p21 association with FTLD in individuals of Western European ancestry. The first study, Finch et al. (2011), confirmed a significant association of rs1990622 (P=0.036) and rs6966915 (P=0.038) with FTLD in a clinically-defined cohort of 640 cases and 822 age- and sex-matched neurologically normal controls (Finch et al., 2011). As with Van Deerlin et al., the major alleles of these SNPs were associated with increased FTLD-TDP risk. The case cohort included 78 *GRN*+ FTLD-TDP individuals and 80 additional patients with confirmed TDP-43 pathology, and individuals with mutations associated with non-TDP-43 pathology were excluded. As was seen in Van Deerlin et al., the associations were more significant when analyzing only *GRN* mutation carriers (P=0.00003 for rs1990622), despite the much smaller sample size (n=78 *GRN*+ individuals). In particular, the percentage of minor (protective) allele homozygotes was drastically lower in *GRN*+ FTLD-TDP (2.6%) than in controls (19.1%) (P=0.003). The authors also showed that among the *GRN*+ individuals, minor allele homozygotes developed FTLD-TDP ~13 years later than major allele homozygotes and heterozygotes (P=0.018) (Finch et al., 2011). This genetic modifier effect was also reported by an independent study in the same year, using a cohort of 50 *GRN*+ FTLD-TDP individuals from four families (Cruchaga et al., 2011). Thus, the minor allele of rs1990622, or a linked variant, appears to protect against developing both *GRN*+ FTLD-TDP and non-Mendelian forms of FTLD-TDP.

The second replication study, van der Zee et al. (2011), replicated the association of rs1990622 with FTLD in cohort of 288 clinically diagnosed FTLD cases who tested negative for *MAPT* or *CHMP2B* mutations, or had autopsy-confirmed TDP-43 pathology (van der Zee et al., 2011). Again, the direction of association matched Van Deerlin et al (2010). This cohort included 13 *GRN*+ FTLD-TDP cases, two individuals with *VCP* mutations, and 14 individuals with autopsy-

confirmed FTLD-TDP. The control cohort consisted of 595 age-matched individuals without family history of neurodegenerative disease and with normal neurological exam function.

A third study attempting to replicate this association was published in 2011 by Rollinson and colleagues (2011). This study used a case cohort of 312 clinically defined FTLD cases, although *GRN* mutation carriers in addition to *MAPT* mutation carriers were excluded, which, based on the results of the previous studies, would likely weaken any association with the *TMEM106B* risk variants (Rollinson et al., 2011a). The control cohort consisted of 248 neurologically normal adults. No significant differences in allele frequencies of rs1990622, rs6966915 or rs1020004 were identified between cases and controls (Rollinson et al., 2011a), although the small sample size, lack of neuropathological homogeneity, and lack of *GRN*+ individuals make a comparison with the other studies difficult.

Finally, a meta-analysis of all four studies, including an additional case-control study from the meta-analysis authors, was published in 2015 (Hernandez et al., 2015). The case-control cohort in this study included 146 clinically-defined FTLD cases and 234 neurologically normal controls from Spain. Individuals with FTLD-causing mutations were not excluded, but only one *MAPT* mutation carrier was present in the cohort. The authors observed a non-significant trend of association (P=0.092) of the major (risk) allele of rs1990622 with increased risk for FTLD in their cohort (Hernandez et al., 2015). By combining their results with the previous four studies, a statistically significant association was seen (P=0.0067), with all studies showing the same direction of effect (Hernandez et al., 2015). However, considerable heterogeneity was seen between studies (P=0.00014), likely due in part to neuropathological heterogeneity present in the clinical cohorts.

Taken together, these data suggest a true association between the *TMEM106B* locus first identified by Van Deerlin et al. (2010) with the neuropathological form of FTLD characterized by TDP-43 inclusions (FTLD-TDP), but not necessarily with other neuropathological forms of FTLD that may be clinically indistinguishable from FTLD-TDP. Moreover, they highlight the importance

of clearly defining a particular trait on the best biological grounds possible (which may not be clinical presentation) when performing association studies.

1.3.1.2 Association of the *TMEM106B* locus with other neurodegenerative phenotypes

Given that many SNP-trait associations identified by GWAS are not replicated in independent studies, the replication of the 7p21 association with FTLD by multiple groups reduces the likelihood that the association is spurious. The extremely high significance (P=1.08x10^{-11} for rs1990622) and effect size (odds ratio=1.64 for rs1990622) of this association further support that this is a true signal. Intriguingly, the importance of genetic variation at 7p21 in neurodegeneration has been further confirmed by several studies that investigated other neurodegenerative and/or neuropathological phenotypes. One study published shortly after the GWAS demonstrated that the allele of rs1990622 associated with increased FTLD-TDP risk also associated with cognitive impairment in amyotrophic lateral sclerosis (ALS) (Vass et al., 2011). Given the clinical, neuropathological, and genetic overlap between FTLD-TDP and ALS (see Chapter 1.1.4.3), this result is intriguing. Furthermore, this result establishes *TMEM106B* genotype as a modifier of clinical presentation in ALS, in addition to FTLD (Cruchaga et al., 2011; Finch et al., 2011; Gallagher et al., 2014).

Additional studies have demonstrated an association of the rs1990622 FTLD-TDP risk allele with increased risk for hippocampal sclerosis of aging, a neurodegenerative condition that presents with similar clinical symptoms as Alzheimer's disease (AD) (Nelson et al., 2015), as well as increased TDP-43 pathology in AD and in neurologically normal individuals (Rutherford et al., 2012; Yu et al., 2015). Two studies reported an association between the rs1990622 risk allele and decreased plasma PGRN levels (Cruchaga et al., 2011; Finch et al., 2011), suggesting that genetic variation at *TMEM106B* may influence disease risk at least partly through an effect on PGRN production and/or secretion. This intriguing possibility has a plausible mechanistic basis, given that loss-of-function *GRN* mutations cause FTLD-TDP (Cruts et al., 2006), and that impaired secretion of PGRN by microglia (which are closely related to peripheral leukocytes) is toxic to neurons (Martens et al., 2012) (see Chapter 1.1.4.2). Finally, the risk allele of rs1990622 has been

associated with decreased neuronal connectivity in the brains of asymptomatic *GRN* mutation carriers (Premi et al., 2014), and reduced temporal lobe gray matter and interhemispheric structures in neurologically normal adults (Adams et al., 2014).

In conclusion, the GWAS sentinel SNP, rs1990622, not only associates with risk of developing FTLD, but also the clinical presentation of FTLD-TDP and ALS, and several phenotypes related to normal and/or pathological brain function. Moreover, although the disease (and non-diseased) states vary across the previously-mentioned studies, one unifying theme may be the presence of TDP-43 pathology, which characterizes FTLD-TDP (both with and without *GRN* mutations) and ALS, and was specifically investigated in the studies of AD and neurologically normal individuals. Thus, genetic variation at the 7p21 locus may specifically influence risk for TDP-43-mediated disease processes.

1.3.2 Increased levels of TMEM106B are implicated in disease

While the association of genetic variants at the *TMEM106B* locus with FTLD-TDP and related phenotypes does not prove that *TMEM106B* is the "causal" gene, this was considered likely to be the case, given that the closest gene 5' of *TMEM106B*, *THSD7A*, is ~380kb away, and the closest gene 3' of *TMEM106B*, *VWDE*, is ~85kb away (Hinrichs et al., 2016). In order to answer this question definitively, however, one would need to functionally dissect the mechanisms underlying the association of *TMEM106B* SNPs with FTLD-TDP risk. This question will be addressed in detail in Chapter 4, in which an unbiased and comprehensive characterization of all candidate causal variants at the *TMEM106B* locus is described, but one result published in the FTLD-TDP GWAS provides a plausible working hypothesis. As discussed in Chapter 1.2, many loci associated with disease by GWAS are also associated with the expression levels of one or more genes (an expression quantitative trait locus (eQTL) effect) (Edwards et al., 2013), and the variants in question are significantly enriched in predicted *cis*-regulatory elements (CREs) (Maurano et al., 2012; Schaub et al., 2012). In Van Deerlin et al. (2010), the FTLD-TDP-associated SNPs were queried for association with gene expression levels using a publicly available eQTL study performed in Epstein-Barr virus (EBV)-transformed B lymphoblastoid cell lines (LCLs) (Dixon

et al., 2007; Van Deerlin et al., 2010). Interestingly, the risk alleles of all three GWAS SNPs were significantly associated with increased mRNA levels of *TMEM106B* in LCLs (P=6.9x10^{-8} for rs1990622), and importantly, these SNPs were not associated with the expression levels of any other genes in the genome (Dixon et al., 2007; Van Deerlin et al., 2010). This result suggests that one or more variants at the *TMEM106B* locus may influence risk for FTLD-TDP by altering *TMEM106B* expression. Additionally, rs1020004, which is in only moderately strong LD with rs1990622 and rs6969915 (r^2=0.7), and which was least strongly associated with FTLD-TDP risk in the GWAS and subsequent replication studies, also showed the weakest association with *TMEM106B* expression levels in this eQTL study (P=1.1x10^{-6}) (Dixon et al., 2007; Van Deerlin et al., 2010). Thus, the association of *TMEM106B* variants with FTLD-TDP risk (and perhaps the other neurodegenerative phenotypes discussed earlier) may be driven by the same functional causal variant(s) underlying the association with *TMEM106B* expression levels. Moreover, these association patterns between *TMEM106B* variants and *TMEM106B* mRNA levels have been replicated in more recent large-scale LCL eQTL studies (Lappalainen et al., 2013; Liang et al., 2013; Stranger et al., 2012), as well as in eQTL studies of T cells, neutrophils, and inflammatory-stimulated monocytes (Fairfax et al., 2014; Peters et al., 2016; Raj et al., 2014).

Since brain is the tissue relevant to FTLD, Van Deerlin and colleagues extracted RNA from 25 prefrontal cortex samples (n=8 *GRN*+ FTLD-TDP, n=10 *GRN*- FTLD-TDP, and n=7 normal) and quantified *TMEM106B* levels using reverse transcription quantitative polymerase chain reaction (RT-qPCR). They then asked whether genotype at rs1990622 associated with *TMEM106B* levels in human brain, and observed that risk allele homozygotes displayed the highest expression levels, heterozygotes displayed intermediate expression levels, and protective allele homozygotes displayed the lowest expression levels (P=0.027) (Van Deerlin et al., 2010). However, the presence of disease was also reported to affect *TMEM106B* levels in this study (see below), and controlling for disease status (*GRN*+ FTLD-TDP, *GRN*- FTLD-TDP or normal) eliminated the association (unpublished data). Subsequent studies also failed to detect genotype-dependent effects on *TMEM106B* levels in brain (Cruchaga et al., 2011; van der Zee et al., 2011). However, these

studies, which consisted of only 23-40 samples, were likely underpowered to detect small effects on expression, as eQTL effects often require hundreds or even thousands of individuals with mixed genotypes in order to be detected (Albert and Kruglyak, 2015). Consistent with this explanation based on sample size, a study of 325 dorsolateral prefrontal cortex samples from aged but neurologically normal individuals revealed a significant association between the rs1990622 risk allele and increased *TMEM106B* mRNA levels (Yu et al., 2015). Thus, *TMEM106B* genotype may increase FTLD risk by increasing *TMEM106B* levels in the brain.

Interestingly, an independent line of evidence suggests that increased *TMEM106B* levels may be involved in FTLD pathogenesis. Van Deerlin and colleagues (2010) also demonstrated that when grouping the same 25 individuals by disease status, individuals with FTLD-TDP had significantly increased *TMEM106B* levels (P=0.045). This effect was driven mostly by the *GRN*+ individuals, who exhibited the highest expression levels (Van Deerlin et al., 2010). This was confirmed in a follow-up study using additional brain regions (temporal cortex and occipital cortex), and increased *TMEM106B* expression was also confirmed at the protein level in *GRN*+ FTLD-TDP frontal cortex samples (Chen-Plotkin et al., 2012). Importantly, this effect appears to be at least partially independent from the effect of *TMEM106B* genotype on *TMEM106B* levels (Van Deerlin et al., 2010), implicating additional regulatory processes that may affect FTLD-TDP pathogenesis by upregulating *TMEM106B*.

Finally, an unrelated study reported that a nonsynonymous protein-coding variant, rs3173615, which is in complete LD with rs1990622, results in altered TMEM106B protein levels, but has no effect at the mRNA level (Nicholson et al., 2013). Indeed, since the reporting of the 7p21 locus association with FTLD-TDP, many papers and reviews have mentioned rs3173615 as a potential causal variant, since it is the only amino acid-altering variant that is in strong LD with the GWAS sentinel SNP (Cruchaga et al., 2011; van der Zee et al., 2011; van der Zee and Van Broeckhoven, 2011). The risk allele of rs3173615 is predicted to change a serine residue to a threonine, which only differ in molecular composition by a single carbon atom. Nicholson and colleagues (2013) cloned the *TMEM106B* open reading frame with either the risk or protective allele

of rs3173615 into an expression construct and transfected these constructs into HeLa and HEK293 cells. They observed no effect of rs3173615 on *TMEM106B* mRNA levels, but observed a ~40% reduction in TMEM106B protein levels in cells transfected with the protective allele construct, in both cell lines (Nicholson et al., 2013). They further show that the protective allele of rs3173615 reduces TMEM106B protein stability, thus explaining the reduced steady-state levels (Nicholson et al., 2013).

While it is possible that rs317615 has a phenotypic effect that is relevant to FTLD risk, there are four concerns regarding this study. First, the *TMEM106B* eQTL effect reported in many studies in different cell and tissue types exists at the mRNA level, and is not addressed in the Nicholson study. Second, other groups have investigated the functional effects of rs3173615 and have failed to detect reproducible effects on cellular phenotypes, including TMEM106B sub-cellular localization, TMEM106B dimerization, and effects on PGRN levels (Brady et al., 2013; Nicholson et al., 2013; Stagi et al., 2014). Finally, and most importantly, this study suffers from a critical technical limitation, in that the antibody used to detect TMEM106B only detects the TMEM106B monomer species in brain. Because of this, the authors assume that TMEM106B dimers, which have been verified in human cell lines and brain by multiple other antibodies (Brady et al., 2013; Chen-Plotkin et al., 2012), do not exist or are not important. I have attempted to replicate these results using our well characterized antibody that detects TMEM106B monomers and dimers (Chen-Plotkin et al., 2012), and have shown that when expressing either TMEM106B isoform under the control of the endogenous *TMEM106B* promoter in HEK293 cells, the decrease in monomer seen in cells expressing the protective allele is compensated for by an increase in the amount of dimer, with no change in total TMEM106B levels (unpublished data). Consistently, a recent study also failed to reproduce an effect of rs3173615 genotype on steady-state TMEM106B levels (Jun et al., 2015). Thus, rs3173615 may affect TMEM106B dimerization kinetics rather than total protein levels. However, further studies are needed to determine whether this variant affects TMEM106B function.

In conclusion, two independent lines of evidence suggest that increased expression of *TMEM106B* may be involved in FTLD pathogenesis. Of particular interest, the association of the FTLD-TDP risk SNPs with *TMEM106B* levels in human cell lines and brain suggests a plausible mechanism by which genetic variation at *TMEM106B* influences risk for FTLD-TDP and other neurodegenerative phenotypes. One variant in complete LD with rs1990622 is predicted to cause a serine (protective allele) to threonine (risk allele) amino acid change, although this variant has not consistently altered any biochemical or cellular phenotypes under investigation.

1.3.3 TMEM106B is involved in the autophagosomal/endolysosomal pathway

At the time of the discovery of the association between the *TMEM106B* SNPs and risk for FTLD-TDP, nothing was known about *TMEM106B*, the sole gene within ~85kb of the FTLD risk SNPs. Within 3 years of the GWAS report, several studies reported initial biochemical and cellular characterizations of TMEM106B. TMEM106B was determined to be a Type II integral transmembrane protein that localizes to late endosomes and lysosomes, as defined by co-localization with the late endosomal/lysosomal markers LAMP-1, Rab7 and Rab9 (Brady et al., 2013; Chen-Plotkin et al., 2012; Lang et al., 2012). This localization pattern has been observed in various murine, non-human primate, and human cell types, including neurons and glia, and has been confirmed in additional studies (Busch et al., 2016; Schwenk et al., 2014; Stagi et al., 2014). In addition, TMEM106B expression has been confirmed in neurons and glia by performing immunohistochemistry on human brain samples from individuals both with and without FTLD-TDP (Busch et al., 2013; Chen-Plotkin et al., 2012).

In light of the fact that *TMEM106B* levels have been implicated in FTLD pathogenesis (see Chapter 1.3.2), many of these studies have overexpressed and/or knocked down *TMEM106B*. Several studies have shown that *TMEM106B* overexpression in various cell types increases lysosomal size, and in some cases affects lysosomal trafficking, although the effects of knockdown are less clear (Brady et al., 2013; Chen-Plotkin et al., 2012; Schwenk et al., 2014; Stagi et al., 2014). Some studies have also reported that overexpression results in impaired lysosomal acidification and protein degradation, as well as cell death, which may be the result of caspase-

and/or lysosome-dependent cell death pathways (Brady et al., 2013; Busch et al., 2016; Chen-Plotkin et al., 2012; Suzuki and Matsuoka, 2016). TMEM106B levels have also been shown to affect neuronal dendritic branching, lysosomal stress responses, and the sensitivity of lysosomes to oxidative damage (Schwenk et al., 2014; Stagi et al., 2014). Importantly, these phenotypes do not seem to be significantly affected by the rs3173615 protein-coding variant, although more rigorous studies are required to definitively rule out potential functions of this SNP (Brady et al., 2013; Stagi et al., 2014). Given that most of the genes known to harbor FTLD-causing mutations have roles in either endolysosomal pathways, which are critical for autophagy (Lamb et al., 2013), or the ubiquitin/proteasome system, the apparent lysosomal function of TMEM106B and its effects on protein degradation are intriguing.

Furthermore, the localization and overexpression phenotypes of *TMEM106B* have been linked to both *GRN* and *C9orf72*, the two most common genes that contain FTLD-causing mutations. First, TMEM106B and PGRN co-localize extensively at LAMP-1+ organelles, and *TMEM106B* overexpression causes increases in intracellular PGRN levels (Brady et al., 2013; Chen-Plotkin et al., 2012; Nicholson et al., 2013). As PGRN has recently been implicated in lysosomal function (Petkau and Leavitt, 2014), *TMEM106B* may thus act as an FTLD risk factor and modifier at least partly through *GRN*-dependent effects on lysosomes. The FTLD-TDP risk allele of *TMEM106B* has also been associated with reduced plasma PGRN levels, and TMEM106B and PGRN levels appear to be negatively correlated in immortalized lymphoblasts and brain (Cruchaga et al., 2011; Finch et al., 2011; Yu et al., 2015). In addition, recent work in my lab has demonstrated that the cellular phenotypes associated with *TMEM106B* overexpression can be rescued or exacerbated by *C9orf72* knockdown and overexpression, respectively (Busch et al., 2016). Consistently, *C9orf72* has recently been implicated in lysosomal processes (Farg et al., 2014; Sellier et al., 2016; Ugolino et al., 2016; Webster et al., 2016; Yang et al., 2016) (see Chapter 1.1.4.3). Thus, the role of *TMEM106B* in FTLD-TDP risk and in modifying FTLD-TDP clinical prognosis in the presence of *GRN* and *C9orf72* mutations may involve interactions between these proteins in lysosomal pathways.

1.3.4 Summary

In summary, common genetic variants at the *TMEM106B* locus have been associated with risk for FTLD-TDP, and this association has been replicated by independent groups. In addition, the risk variants at *TMEM106B* also appear to modify age at death and/or age at onset in FTLD-TDP patients with disease-causing *GRN* or *C9orf72* mutations. Finally, *TMEM106B* risk variants influence risk for hippocampal sclerosis of aging, TDP-43 pathology, cognitive impairment in ALS, PGRN levels, temporal lobe gray matter volume, brain interhemispheric structure volume, and functional network connectivity in asymptomatic *GRN* mutation carriers. Thus, genetic variation at *TMEM106B* appears to play an important role in many phenotypes related to neurodegeneration, and possibly normal brain function, many of which are associated with TDP-43 pathology.

Two independent lines of evidence suggest that *TMEM106B* may be the causal gene that affects these phenotypes in a genotype-dependent manner. First, genome-wide eQTL studies performed in several white blood cell types have revealed a significant association between the *TMEM106B* risk SNP alleles and increased *TMEM106B* expression levels, and this association has been confirmed to exist in human prefrontal cortex. Importantly, these genome-wide studies fail to show a significant association between FTLD-TDP-associated SNPs and the expression levels of any other genes. In addition, TMEM106B levels are increased in FTLD-TDP brains compared to healthy controls, irrespective of *TMEM106B* genotype. Consistent with a potentially pathogenic role of increased TMEM106B levels, cell-based studies have repeatedly shown that *TMEM106B* overexpression results in enlarged, poorly acidified lysosomes, impaired protein degradation, and cell death. Finally, several lines of evidence suggest that *TMEM106B* function may be linked to those of *GRN* and *C9orf72* in lysosomal pathways, potentially explaining the ability of *TMEM106B* genotype to modify disease course in FTLD patients with disease-causing mutations in these genes.

Taken together, these results implicate *TMEM106B* as an important genetic risk factor for FTLD and other neurodegenerative phenotypes. Given our poor current understanding of the pathogenesis of FTLD and related conditions, and the lack of effective treatments for these

devastating diseases, *TMEM106B* provides a unique opportunity to improve our understanding of the mechanisms underlying this disease.

CHAPTER 2: TMEM106B, THE RISK GENE FOR FRONTOTEMPORAL DEMENTIA, IS REGULATED BY THE MICRORNA-132/212 CLUSTER AND AFFECTS PROGRANULIN PATHWAYS

by

Alice S. Chen-Plotkin, Travis L. Unger, Michael D. Gallagher, Emily Bill, Linda K. Kwong, Laura Volpicelli-Daley, Johanna I. Busch, Sebastian Akle, Murray Grossman, Vivianna Van Deerlin, John Q. Trojanowski, and Virginia M.-Y. Lee

This chapter was written by Alice Chen-Plotkin, with input from Michael Gallagher and other authors, and originally published in the *Journal of Neuroscience* (August 15, 2012, Volume 32, Issue 33, Pages 11213-27). The majority of the microRNA work, shown in Figures 2.3 and 2.4, was performed by Michael Gallagher, and relates to the rest of the dissertation by implicating increased TMEM106B levels in FTLD pathogenesis.

2.1 ABSTRACT

Frontotemporal lobar degeneration with TDP-43 inclusions (FTLD-TDP) is a fatal neurodegenerative disease with no available treatments. Mutations in the progranulin gene (*GRN*) causing impaired production or secretion of progranulin are a common Mendelian cause of FTLD-TDP; additionally, common variants at chromosome 7p21 in the uncharacterized gene *TMEM106B* were recently linked by genome-wide association to FTLD-TDP with and without *GRN* mutations. Here we show that *TMEM106B* is neuronally-expressed in postmortem human brain tissue, and that expression levels are increased in FTLD-TDP brain. Furthermore, using an unbiased, microarray-based screen of over 800 microRNAs, we identify microRNA-132 as the top microRNA differentiating FTLD-TDP and control brains, with <50% normal expression levels of three members of the microRNA-132 cluster (microRNA-132, microRNA-132*, and microRNA-212) in disease. Computational analyses, corroborated empirically, demonstrate that the top mRNA target of both microRNA-132 and microRNA-212 is *TMEM106B*; both microRNAs repress TMEM106B expression through shared microRNA-132/212 binding sites in the *TMEM106B* 3'UTR. Increasing TMEM106B expression to model disease results in enlargement and poor acidification of endo-lysosomes, as well as impairment of mannose-6-phosphate-receptor trafficking. Finally, endogenous neuronal TMEM106B co-localizes with progranulin in late endo-lysosomes, and TMEM106B over-expression increases intracellular levels of progranulin. Thus, *TMEM106B* is an FTLD-TDP risk gene, with microRNA-132/212 depression as an event which can lead to aberrant over-expression of TMEM106B, which in turn alters progranulin pathways. Evidence for this pathogenic cascade includes the striking convergence of two independent, genomic-scale screens on a microRNA:mRNA regulatory pair. Our findings open novel directions for elucidating miRNA-based therapies in FTLD-TDP.

2.2 INTRODUCTION

The neurodegenerative dementia FTLD-TDP is a sporadic and familial neurodegenerative disease causing progressive impairment in language, behavioral control, or both (Baker et al., 2006; McKhann et al., 2001). One of the major forms of presenile dementia (Ratnavalli et al., 2002), FTLD-TDP is characterized by ubiquitinated inclusions comprised primarily of the HIV TAR-DNA binding protein of 43kD, or TDP-43 (Arai et al., 2006; Neumann et al., 2006). These TDP-43 inclusions are also found in the motor neuron disease amyotrophic lateral sclerosis (ALS) (Arai et al., 2006; Neumann et al., 2006). Mutations in the TDP-43 gene (*TARDBP*) are rare genetic causes of both FTLD-TDP (Benajiba et al., 2009) and ALS (Gitcho et al., 2008), suggesting that FTLD-TDP and ALS might be two phenotypic ends of one disease spectrum (Chen-Plotkin et al., 2010a). Additionally, mutations in the progranulin gene (*GRN*) – a secreted growth factor (He and Bateman, 2003) – are a major Mendelian cause of FTLD-TDP (Baker et al., 2006; Cruts et al., 2006) and account for ~10% of all cases (Gass et al., 2006). Deficiency of progranulin production (Baker et al., 2006; Cruts et al., 2006) or secretion (Mukherjee et al., 2008; Shankaran et al., 2008) appears to be the disease mechanism in these autosomal dominant FTLD-TDP cases (Cruts and Van Broeckhoven, 2008).

To identify additional risk factors for FTLD-TDP, we previously conducted a genome-wide association study (GWAS) and showed that chromosome 7p21 variants within the gene *TMEM106B* confer increased risk of FTLD-TDP, with an odds ratio of 1.6 (Van Deerlin et al., 2010), and this association has been replicated (van der Zee et al., 2011). Intriguingly, decreased plasma progranulin levels correlate with *TMEM106B* risk genotypes (Finch et al., 2011), and, in ALS patients, *TMEM106B* genotypes associated with FTLD-TDP increase the risk of developing dementia (Vass et al., 2011). While these observations correlate with *TMEM106B* genotype, they do not provide mechanistic evidence that *TMEM106B* is the causative 7p21 genetic signal observed in the GWAS. Furthermore, very little is known about TMEM106B, a 274 amino-acid, predicted single transmembrane domain protein, with no yeast orthologue and homology only to two other uncharacterized members of the TMEM106 family.

Here, we investigate the genetic regulation and pathophysiological function of TMEM106B, both of which were previously unknown. We demonstrate that TMEM106B is elevated in FTLD-TDP brains. We further show that TMEM106B is normally repressed by microRNA-132 and microRNA-212, which are significantly decreased in FTLD-TDP. Finally, we demonstrate that TMEM106B over-expression in turn disrupts endosomal-lysosomal pathways, sequesters progranulin in TMEM106B positive late endosomes or lysosomes, and increases intracellular levels of progranulin. We thus establish *TMEM106B* mechanistically as the 7p21 genetic risk factor for FTLD-TDP and elucidate pathophysiological steps which may be amenable to targeted intervention in an otherwise fatal disease.

2.3 RESULTS

2.3.1 TMEM106B shows increased expression in FTLD-TDP brain

TMEM106B is a minimally characterized protein. Thus, we first sought to investigate its expression in normal and FTLD-TDP brain tissue. We thus raised an affinity-purified polyclonal antibody, N2077, which specifically recognizes an N-terminus peptide sequence unique to TMEM106B.

In both HEK293 cells (**Figure 2.1A**) and murine primary cortical neurons (**Figure 2.1B**) transfected with a FLAG-tagged TMEM106B construct, staining patterns for N2077 and anti-FLAG antibody showed nearly perfect overlap. Similarly, immunoblots performed on sequentially-extracted HEK293 cell lysates transfected with FLAG-tagged TMEM106B constructs showed the same band when probed with N2077 or antibodies against the tag (**Figure 2.1C**), again demonstrating that the N2077 antibody recognizes TMEM106B. Furthermore, both N2077 and the anti-FLAG antibodies recognize TMEM106B in RIPA-extracted samples, consistent with the predicted transmembrane character of TMEM106B. Finally, preabsorption of N2077 with the peptide immunogen resulted in the disappearance of TMEM106B immunobands on immunoblotting (**Figure 2.1D**), as well as the disappearance of TMEM106B staining within cellular structures on immunofluorescence (data not shown). Taken together, these results demonstrate the specificity of the N2077 antibody for TMEM106B in both biochemical and cell biological contexts.

Unexpectedly, however, the molecular weight of the main species recognized by both N2077 and antibodies directed against the FLAG tag was approximately 75kD, with a fainter band occasionally seen at 40kD, while the predicted molecular weight of TMEM106B is 31kD. We accordingly tested other commercially available antibodies against TMEM106B; only one recognized immunobands that increased with over-expression of TMEM106B, and this antibody demonstrated the same 75kD and 40kD bands (**Figure 2.1D**), with possibly greater affinity for the 40kD band than N2077. To investigate the apparent discrepancy between predicted and observed molecular weights of TMEM106B, we manipulated the protein in various ways. Dephosphorylating or reducing cell lysates with DTT did not change the molecular weight (**Figure 2.1E**). Incubating

the protein at 37°C or above, however, demonstrated that the 75kD species recognized by both N2077 and antibodies against the TMEM106B construct tag was very sensitive to heat, even in the presence of protease inhibitors. Specifically, when cell lysates containing TMEM106B were kept at 37°C for 30 minutes, several changes were observed in the immunoblot pattern (**Figure 2.1F**). First, more than 50% of the protein was lost. Second, the 40kD band became more prominent. Third, when cell lysates containing TMEM106B were heated to 56°C or above, both the 75kD and 40kD bands were lost, and a high-molecular weight species was seen at the top of the gel, again consistent with the transmembrane character of TMEM106B. Human brain homogenates behaved similarly, with the exception that the 40kD band was much less prominent (**Figure 2.1G**).

Finally, TMEM106B has recently been reported to be glycosylated, with at least five glycosylation sites C-terminal to the predicted transmembrane domain (Lang et al., 2012). We confirmed this finding and extended it to brain tissue by deglycosylating cell lysates and human brain homogenates. Specifically, we found that if we combined treatment with N-glycosidase F with short pre-treatment with heat, the 40kD band (prominent in cell lysates and visible although less prominent in human brain homogenates) collapsed to the expected 31kD weight of TMEM106B, while the 75kD band collapsed to ~60kD (**Figure 2.1H**). Taken together, these data suggest that TMEM106B is glycosylated, detergent-soluble, and exists primarily in a heat-sensitive protein complex migrating at an apparent molecular weight of 75kD. Accordingly, all subsequent biochemical experiments performed for quantitation were done with the samples kept at 4°C.

Having demonstrated the specificity of our N2077 antibody, we used it to evaluate the expression of TMEM106B in human frontal cortex brain tissue (n=5 each of normal cases. FTLD-TDP without *GRN* mutations, FTLD-TDP with *GRN* mutations). TMEM106B was abundantly expressed in human brain tissue, with a cytoplasmic perikaryal localization in neurons from normal individuals (**Figure 2.2A**), reminiscent of TMEM106B's expression pattern in murine primary cortical neurons (**Figure 2.1B**). Intriguingly, TMEM106B appeared more widely distributed in the cytoplasm of neurons from individuals with FTLD-TDP, and particularly FTLD-TDP with *GRN* mutations (*GRN*(+) FTLD-TDP, **Figure 2.2A**). In these individuals, TMEM106B expression

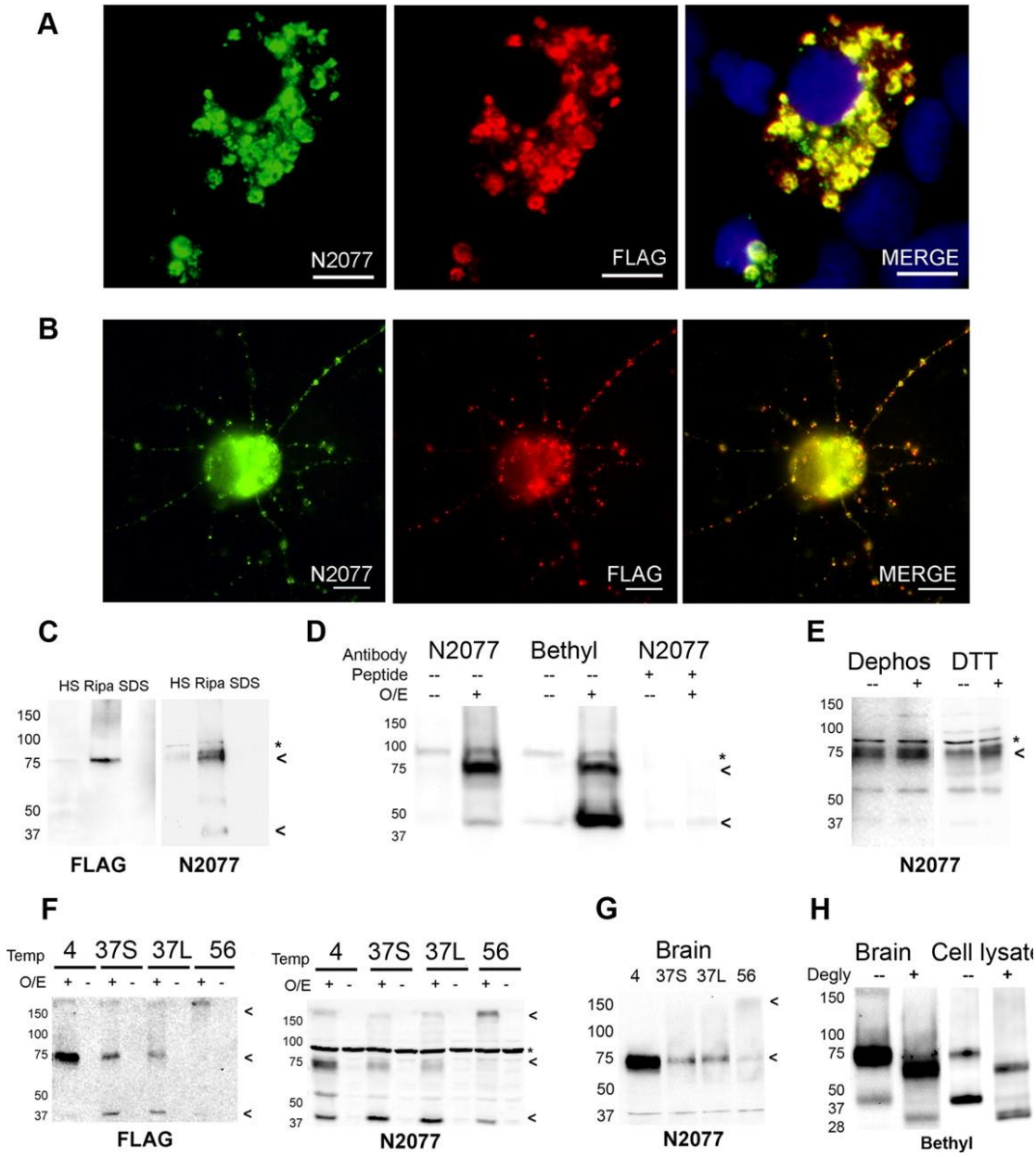


Figure 2.1. TMEM106B antibody and protein characterization.
We raised a novel polyclonal antibody (N2077) recognizing amino acids 4-19 of TMEM106B.
(A,B) HEK293 cells (Panel A) or murine primary cortical neurons (Panel B) were transfected with FLAG-tagged TMEM106B. Double label immunofluorescence microscopy demonstrates that anti-FLAG antibody recognizes the same cellular structures as anti-TMEM106B antibody (N2077), demonstrating the specificity of our antibody in a cell biological context.
(C) HEK293 cells transfected with FLAG-tagged TMEM106B were sequential extracted into high-salt (HS), RIPA, and 2% SDS (SDS) buffers. A RIPA-soluble 75kD band was recognized by both anti-FLAG antibody and anti-TMEM106B antibody (N2077), demonstrating the specificity of our antibody in a biochemical context.
(D) TMEM106B over-expression in HEK293 cells resulted in the appearance of bands at 75kD and 40kD, detected by both N2077 (first two columns) and a commercially available N-terminus

antibody (Bethyl, middle two columns). Both bands disappeared when N2077 was preabsorbed with immunizing peptide (last two columns). O/E indicates whether TMEM106B was over-expressed (+) or not (-). Peptide indicates whether the antibody was preabsorbed with peptide immunogen (+) or not (-).
(E) Neither dephosphorylation with lambda phosphatase, nor treatment with the reducing agent DTT changed the electrophoretic mobility of TMEM106B. Dephos = dephosphorylated, DTT = DTT treated; (-) indicates control, (+) indicates treated condition.
(F,G) TMEM106B shows unusual heat sensitivity at temperatures above 4°C, even in the presence of protease inhibitors. In cell lysates from HEK293 cells transfected with FLAG-tagged TMEM106B (Panel F), TMEM106B appeared primarily as a 75kD band when samples were kept on ice (4). When heated at 37°C for 30 minutes (37S) or 45 minutes (37L), the 75kD band faded, and a 40kD species began to appear. At higher temperatures (e.g. 56°C for 15min, lane labeled 56), both the 40kD and 75kD bands were lost and a >150kD aggregate appeared at the top of the gel. Overexpression (O/E) indicates whether TMEM106B was over-expressed (+) or not (-). Blots were probed with both anti-FLAG and N2077 antibodies, demonstrating the specificity of the bands. Similar heat-sensitivity was seen for TMEM106B extracted from human brain tissue from normal individuals (Panel G). In contrast to HEK293 cell lysates, however, the 40kD band was never prominent in human brain samples. For all panels, samples were extracted into RIPA buffer, and equal amounts of protein were loaded into all lanes.
(H) TMEM106B from human brain samples (left), or over-expressed in HEK293 cells (right) was deglycosylated with PNGase F after short pre-treatment at 37°C. The 75kD and 40kD bands observed prior to deglycosylation (-) collapsed to lower molecular weight species of ~60kD and 31kD, respectively, after PNGase F treatment (+). Note that the predicted molecular weight of TMEM106B is 31kD. Blots probed with the Bethyl TMEM106B antibody.
All immunoblots: Arrowheads indicate TMEM106B species. *non-specific band.

extended beyond the cell body into neuronal processes, filling the cytoplasm of the cells. Of note, none of the commercially available TMEM106B antibodies produced a clear staining pattern in human brain tissue (data not shown), whereas N2077's perikaryal neuronal staining was robust and disappeared upon preabsorption of the antibody with the peptide immunogen (**Figure 2.2B**).

These immunohistochemical findings, as well as preliminary data reported in our prior GWAS (Van Deerlin et al., 2010), suggested that TMEM106B expression is increased in FTLD-TDP. We therefore quantified TMEM106B expression in FTLD-TDP brain.

At the mRNA level, quantitative reverse-transcription PCR (QRT-PCR) demonstrated that *TMEM106B* brain expression was lowest in normal individuals (n=6), higher in FTLD-TDP patients without *GRN* mutations (*GRN*(-) FTLD-TDP, n=7) and highest in *GRN*(+) FTLD-TDP (n=5, see **Table 2.1** for patient and control characteristics). These trends were observed in multiple areas of the brain (**Figure 2.2C**), with the most striking differences seen in frontal cortex, corroborating our previously published report (Van Deerlin et al., 2010) with a wider range of samples. For a subset of patients, never-thawed frontal cortex samples were available for protein quantitation, and

quantitation of immunoblots performed on these frontal cortex samples corroborated the mRNA finding that TMEM106B expression is increased in *GRN*(+) FTLD-TDP (**Figure 2.2D**). It is worth noting that human brain TMEM106B expression appears to be much higher than in HEK293 cells, which show negligible expression by immunoblot in the absence of over-expression. In addition, we have previously shown that the *GRN*(+) FTLD-TDP brains used in this study do not have more severe TDP-43 pathology than the *GRN*(-) FTLD-TDP brains (Chen-Plotkin et al., 2010b), suggesting that general mechanisms of TDP-43 mediated pathology and/or neurodegeneration cannot account for the observed differences in TMEM106B expression.

Thus, TMEM106B appears to be abundantly expressed in human brain tissue, with particularly high levels of expression in *GRN*(+) FTLD-TDP brain, and possible modest increases in *GRN*(-) FTLD-TDP brain as well.

2.3.2 The microRNA-132 cluster is decreased in FTLD-TDP brain

We next sought to identify genetic modifiers of TMEM106B expression which might be responsible for the observed increase in expression in FTLD-TDP. To do this, and to simultaneously pursue effectors of a distinct mRNA expression signature previously demonstrated in FTLD-TDP (Chen-Plotkin et al., 2008), we performed global microRNA (miRNA) expression profiling. MiRNAs – small noncoding RNAs of ~22bp – have emerged in recent years as major regulators of gene expression, binding to complementary sequences in the 3' untranslated region (3'UTR) of specific mRNAs and repressing expression by targeting mRNAs for degradation or blocking translation (Filipowicz et al., 2008). In addition, miRNA dysfunction has been implicated in a growing number of diseases, including the neurodegenerative disorders Alzheimer's disease (Hebert et al., 2008; Wang et al., 2008), Parkinson's disease (Kim et al., 2007a), and spinocerebellar ataxia (Lee et al., 2008).

Utilizing a microarray platform testing all known human miRNAs (n=836), we compared frontal cortex tissue from FTLD-TDP patients (n=12) and controls (n=6, see **Table 2.1** for patient and control characteristics). Three miRNAs in the miR-132 cluster – miR-132, miR-132*, and miR-212 – all emerged within the 11 miRNAs showing significantly dysregulated expression in FTLD-

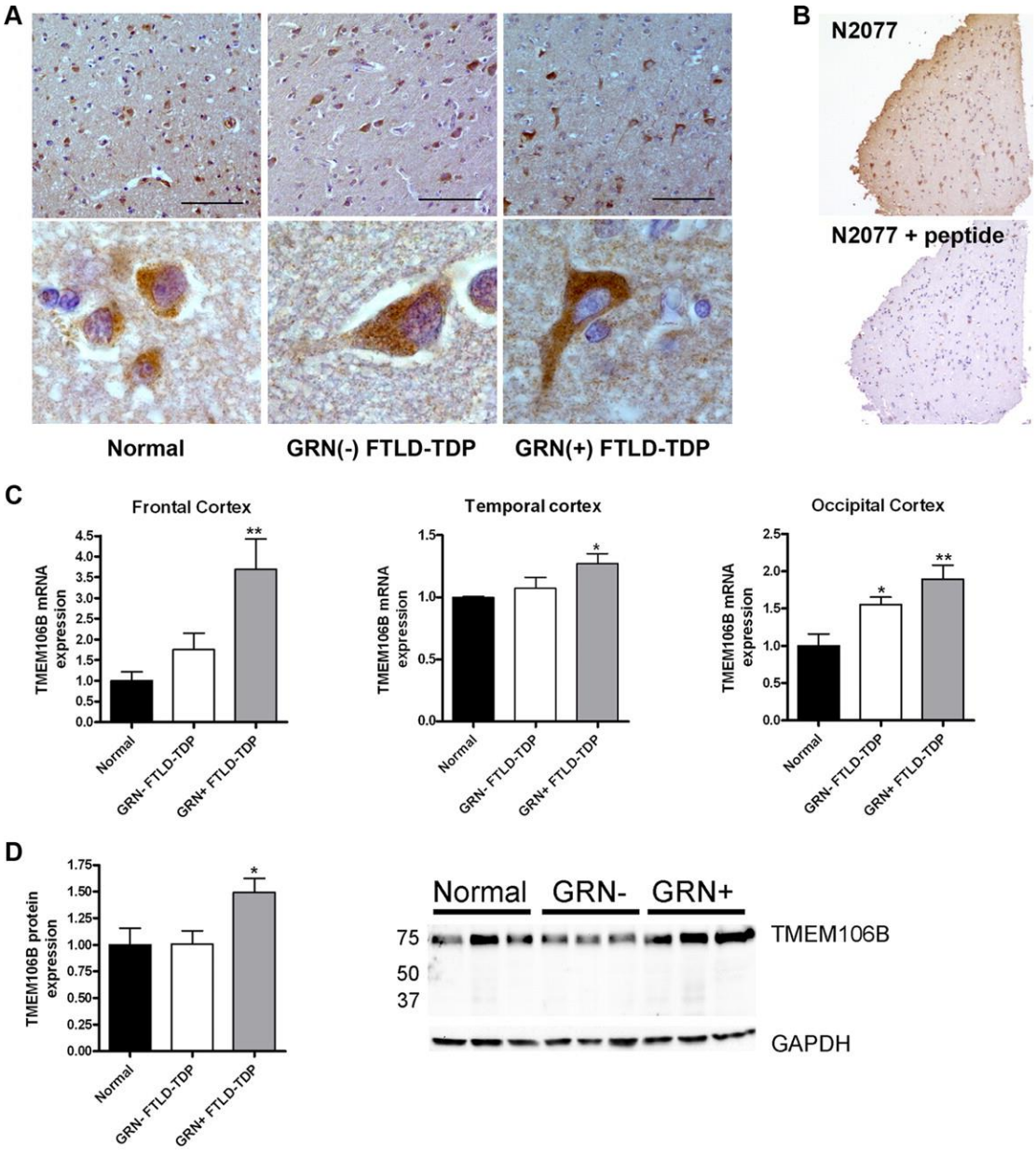


Figure 2.2. *TMEM106B* expression is increased in FTLD-TDP.
(A) Immunohistochemical staining was performed with N2077 anti-TMEM106B antibody on frontal cortex brain sections from age-matched controls (Normal), *GRN*(-) FTLD-TDP and *GRN*(+) FTLD-TDP). *GRN(+)* FTLD-TDP patients had more diffuse TMEM106B staining, extending throughout the cell body and into neuronal processes. Representative lower-magnification images (top) and higher-magnification images of typical neurons from the same field (bottom) are shown. Scale bar represents 100µm.
(B) Neuronal staining (top) was abolished with preabsorption of N2077 with the immunizing peptide (bottom). Scale bar represents 200µm.
(C) Total mRNA was isolated from neurologically normal controls (n=6), *GRN(-)* FTLD-TDP (n=7), and *GRN(+)* FTLD-TDP (n=5), and *TMEM106B* transcript expression was measured by QRT-PCR in multiple brain regions. Compared to both normal controls and to *GRN(-)* FTLD-TDP, *GRN(+)*

FTLD-TDP had significantly higher levels of *TMEM106B* expression. Means +/- SEM shown. *p<0.05, **p<0.01
(D) Frontal cortex protein was RIPA-extracted. Equal amounts of total protein from neurologically normal controls (n=4), *GRN*(-) FTLD-TDP (n=3), and *GRN*(+) FTLD-TDP (n=4) were loaded, and immunoblots were probed for TMEM106B. Corroborating our mRNA findings, *GRN*(+) FTLD-TDP brain showed higher levels of TMEM106B protein expression. Quantification (means +/- SEM) includes all available samples; representative subset immunoblot is also shown.

Group	***N***	***Gender***	***Age at death Median (IQR)***
FTLD-TDP with *GRN* mutations c.26C>A (A9D) c.911G>A (W304X) c.1252C>T (R418X) – 2 cases c.1477C>T (R493X)	5	2M/3F	68 (65-76)
	7	3M/4F	68 (56-73)
FTLD-TDP without *GRN* mutations	6	4M/2F	71 (60-75)
Neurologically normal controls			

Table 2.1. Human brain samples.
Characteristics of postmortem brain samples used for this study. All *GRN* genetic variants used in this study are believed to be pathogenic (http://www.molgen.ua.ac.be/admutations/). M=Male, F=Female
Nomenclature follows cDNA sequence NM_002087.2.

TDP (**Figure 2.3A**). A fourth putative member of the miR-132 cluster – miR-212* – is not in the microRNA database miRBase 17 (http://www.mirbase.org/index.shtml) and was not represented on the microarray.

The top miRNA thus identified was miR-132 (p=0.0001, **Figure 2.3A**), a CREB-activated miRNA previously reported to play diverse roles in neuronal differentiation (Magill et al., 2010; Vo et al., 2005). In addition, by corroborative QRT-PCR, miR-132, miR-132*, and miR-212 all showed <50% expression in both *GRN*(-) FTLD-TDP and *GRN*(+) FTLD-TDP compared to normal controls (QRT-PCR, **Figure 2.3B**). This decreased expression relative to normal controls persisted in both *GRN*(-) FTLD-TDP and *GRN*(+) FTLD-TDP subgroups, removing the possibility that one subgroup is driving the effect; a significant difference also persisted when quantitation was normalized to the

brain-expressed miRNA miR-124, removing the possibility that neuronal loss associated with FTLD-TDP is driving the effect (data not shown). However, absolute levels of miR-132 were ~100 times higher than miR-132* and miR-212 in all groups (**Figure 2.3B**).

MiR-132, miR-132*, and miR-212 are all generated from a shared primary miRNA transcript (pri-miRNA) on chromosome 17; this pri-miRNA is then processed into 2 pre-miRNAs, which in turn become the three mature miRNAs (**Figure 2.3C**). Interestingly, miR-132, miR-132*, and miR-212 levels were highly correlated on an individual-to-individual basis (R^2 values > 90%, **Figure 2.3D**), suggesting that the observed decrease of all three miRNAs may result from decreased expression of the pri-miRNA transcript in FTLD-TDP.

2.3.3 MiR-132 and miR-212 are dual repressors of TMEM106B through shared binding sites in the 3'UTR

Having identified the miR-132 cluster by genomic-scale screening for FTLD-TDP-dysregulated miRNAs, we next performed computational analyses of miR-132 cluster targets using the program TargetScan (Lewis et al., 2003) (release 5.1, http://www.targetscan.org). Strikingly, the top computationally-predicted target of both miRNA-132 and miR-212 was *TMEM106B*, ranked 1 of 283 predicted mRNA targets (**Table 2.2**). Specifically, miR-132 and miR-212 share one recognition sequence, and two sites within the *TMEM106B* 3'UTR contain this "seed match" (**Figure 2.4A**, nucleotides in red). Indeed, on an individual-by-individual basis, irrespective of disease status, frontal cortex levels of both miR-132 and miR-212 showed significant inverse associations with frontal cortex levels of *TMEM106B* (linear regression one-tailed p-value = 0.03 for miR-132, 0.05 for miR-212).

To determine if the computationally predicted relationship between miR-132/212 and *TMEM106B* could be demonstrated experimentally, we transfected each miRNA into HEK293 cells. Transfection of either miR-132 or miR-212 resulted in a 40% decrease of endogenous mRNA levels of *TMEM106B*, suggesting that these miRNAs are *bona fide* negative regulators of *TMEM106B* (**Figure 2.4B**).

A

Rank	MicroRNA	P-value	Fold change (FTLD/control)
1	Hsa-miR-132	0.0001	0.5529
2	Hsa-miR-132*	0.0011	0.6662
3	Hsa-miR-557	0.0108	1.7440
4	Hsa-miR-876-5p	0.0124	0.6806
5	Hsa-miR-193a-3p	0.0131	1.3188
6	Hsa-miR-212	0.0180	0.8073
7	Hsa-miR-1308	0.0210	0.6694
8	Hsa-miR-92b	0.0229	1.1024
9	Hsa-miR-138	0.0385	0.6893
10	Hsa-miR-1826	0.0424	0.8552
11	Hsa-miR-190	0.0447	1.2299

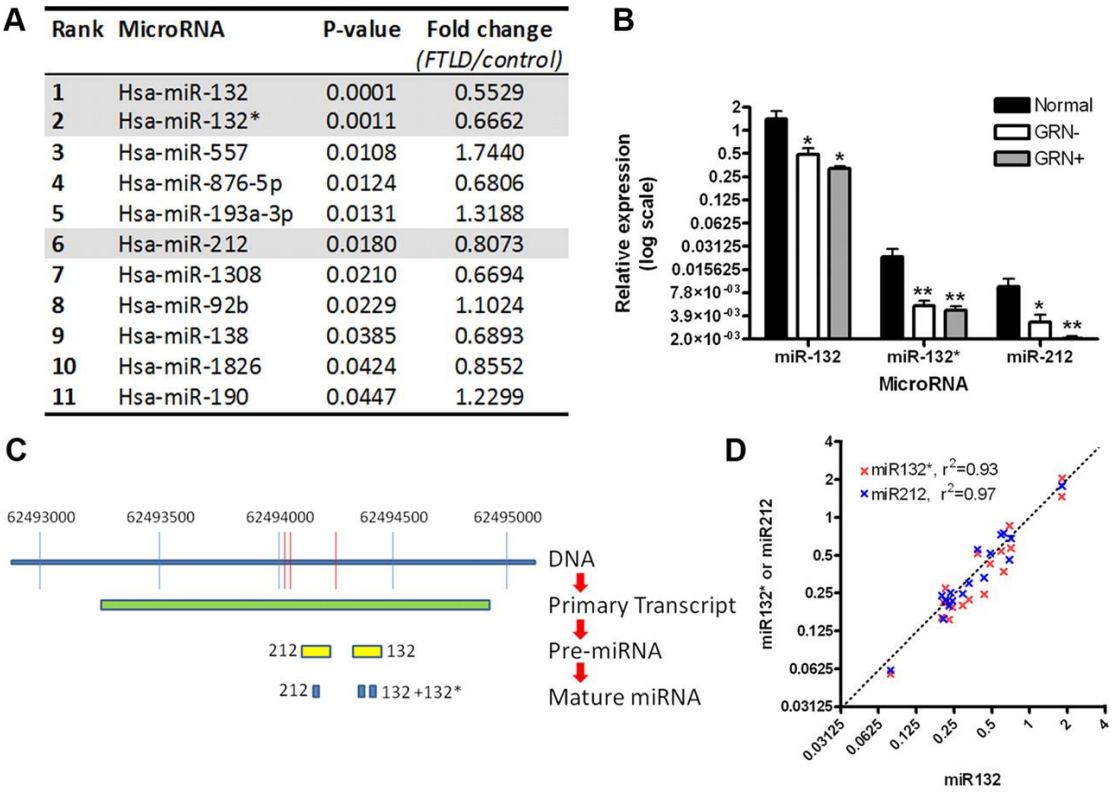


Figure 2.3. The microRNA 132/212 cluster is decreased in FTLD-TDP.
(A) Frontal cortex samples from normal controls (n=6) and patients with FTLD-TDP (n=12) were evaluated by microarray for differentially expressed microRNAs. Eleven microRNAs demonstrated nominally significant differences in FTLD-TDP compared to controls, with microRNA-132 (miR-132) demonstrating the most significant association (p=0.0001), and a decrease of ~50% in FTLD-TDP.
(B) Confirmation of microarray screening results by QRT-PCR showed significant 65-75% reductions in miR-132, miR-132*, and miR-212 for FTLD-TDP frontal cortex samples from patients with (GRN+, n=5) and without (GRN-, n=7) progranulin mutations, compared to neurologically normal controls (n=6). Absolute levels of miR-132, however, were 50-100 times higher than levels of either miR-132* or miR-212. Relative expression (means +/- SEM) of microRNAs are shown on a log2 scale, calibrated to one normal sample's miR-132 expression value. *p<0.05, **p<0.01.
(C) Three of the 11 microRNAs with differential expression in FTLD-TDP are known to arise from processing of the same CREB-responsive primary transcript. Specifically, a primary transcript on chromosome 17 gives rise to pre-microRNA (pre-miRNA) 132 and pre-miR-212. Pre-miR-132 is then further processed to yield mature miR-132 and miR-132*, while pre-miR-212 yields mature mir-212.
(D) MiR-132 (x-axis), miR-132* (y-axis, red), and miR-212 (y-axis, blue) show highly correlated expression levels across all 18 human samples, suggesting that the observed decrease of all 3 microRNAs may be due to decreased expression of the shared primary transcript. For each miRNA, values are normalized to one sample within that group to account for the much higher expression levels of mir-132.

Rank	Target Gene	Gene Name	Total context score
1	*TMEM106B*	transmembrane protein 106B	-0.96
2	*ZNF516*	zinc finger protein 516	-0.89
3	*LEMD3*	LEM domain containing 3	-0.81
4	*CDC2L6*	cell division cycle 2-like 6 (CDK8-like)	-0.81
5	*GMFB*	glia maturation factor, beta	-0.79
6	*ETNK1*	ethanolamine kinase 1	-0.79
7	*TIMM9*	translocase of inner mitochondrial membrane 9 homolog	-0.75
8	*SLC26A7*	solute carrier family 26, member 7	-0.71
9	*CCDC88A*	coiled-coil domain containing 88A	-0.69
10	*MIA3*	melanoma inhibitory activity family, member 3	-0.68

Table 2.2. Top ten mRNA targets of miR-132/212, as predicted by TargetScan.
The total context score reflects the likelihood of true targeting and is related to the number of predicted binding sites, the quality of each binding site, and the genomic context. A more negative value suggests a stronger probability of true targeting. Of note, miR-132 and miR-212 have the same target binding sequence, thus accounting for the shared mRNA targets.

We next evaluated the site-specific contributions of the two predicted miR-132/212 binding sites within the *TMEM106B* 3'UTR (Sites 1 and 2 in **Figure 2.4A**) on miR-132/212 regulation of *TMEM106B*. First, we used a "subtraction" strategy – removing miR-132/212 binding sites from the native *TMEM106B* 3'UTR by truncations and by site-directed deletions (**Figure 2.4A**). We expressed these *TMEM106B* 3'UTR mutants in HEK293 cells and assayed their expression levels under the regulation of endogenous miR-132/212.

Expression of a *TMEM106B* construct lacking the predicted 3'UTR miR-132/212 binding sites resulted in high TMEM106B protein expression (**Figure 2.4C**). Inclusion of the 3' end of the 'UTR in the full 3'UTR-containing *TMEM106B* construct resulted in a 65% decrease of TMEM106B protein (**Figure 2.4C**) and mRNA levels (**Figure 2.4D**), compared to the construct without miR-132/212 binding sites – demonstrating the repressive function of this portion of the 3'UTR.

Moreover, a targeted deletion eliminating only the seed region of miR-132/212 Site 1 from the full 3'UTR construct resulted in significant restorations of TMEM106B protein (**Figure 2.4C**) and mRNA (**Figure 2.4D**) levels, while a targeted deletion eliminating only the seed region of miR-132/212 Site 2 resulted in negligible change in TMEM106B levels. Finally, a construct containing targeted deletions of both miR-132/212 Site 1 and miR-132/212 Site 2 seed regions resulted in the largest restorations (>50%) of TMEM106B expression at the protein (**Figure 2.4C**) and mRNA (**Figure 2.4D**) levels. Thus, miR-132/212 Site 1 and miR-132/212 Site 2 are necessary in mediating the repression of *TMEM106B* by endogenous microRNAs, with greater repressor activity from Site 1.

Second, we employed an "addition" strategy – selectively adding the full miR-132/212 binding Site 1 or 2 to luciferase reporter constructs in the 3' regulatory region. Co-transfection of luciferase reporter constructs with miR-132 or miR-212 into HEK293 cells resulted in a significant reduction in luciferase reporter activity for both the Site 1 construct (**Figure 2.4E**) and the Site 2 construct (**Figure 2.4F**). This reduction was not seen when cells were treated with an irrelevant microRNA or with no microRNAs, or when mir-132/212-site-containing reporters were substituted with either empty vector (**Figures 2.4E and 2.4F**, grey bars) or a luciferase reporter construct inserting the same base pairs in a scrambled order that would abolish microRNA binding (**Figures 2.4E and 2.4F**, black bars). Thus, either miR-132/212 Binding Site 1 or miR-132/212 Binding Site 2 of the *TMEM106B* 3'UTR is sufficient to confer regulatory activity by miR-132 and miR-212.

We next evaluated whether miR-132/212 activation by a physiologic stimulus in a neuronal cell would demonstrate the same regulatory effect of miR-132/212 on *TMEM106B*. Previous studies have shown that in neurons, miR-132 and miR-212 are CREB-activated miRNAs (Impey et al., 2004; Magill et al., 2010) and that treatment with BDNF, which activates CREB-mediated transcription, also increases miR-132 and miR-212 expression. Because the miR-132/212 binding sites in the *TMEM106B* 3'UTR are not conserved beyond primates, we used retinoic acid induced neuronally-differentiated human SHSY5Y cells. BDNF treatment in SHSY5Y cells resulted in a rapid (1-2hr) increase in both the known CREB-responsive gene *fos* and precursor forms of miR-

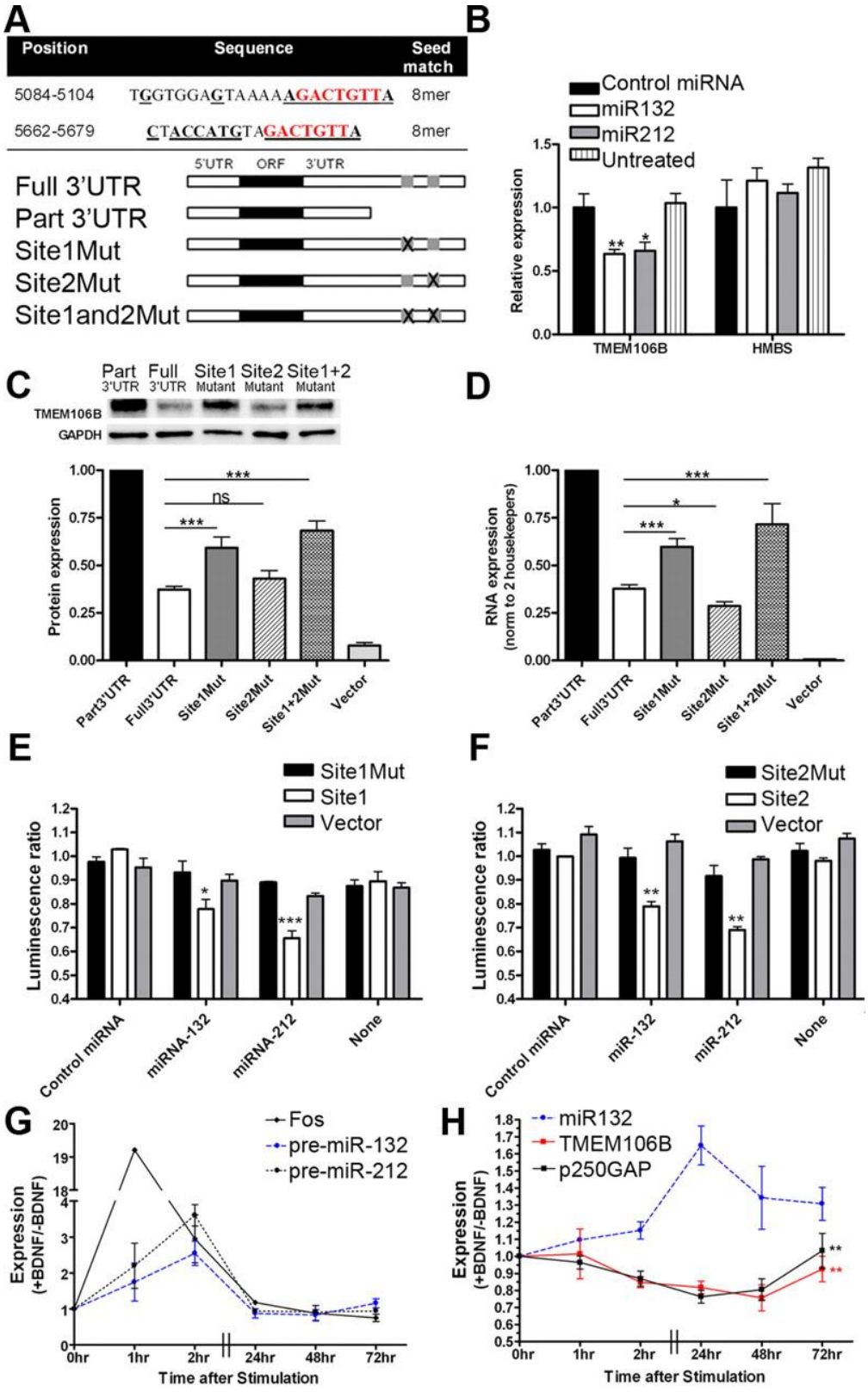

A
Position
Sequence
Seed match
5084-5104
TGGTGGAGTAAAAAGACTGTTA
8mer
5662-5679
CTACCATGTAGACTGTTA
8mer
5'UTR
ORF
3'UTR
Full 3'UTR
Part 3'UTR
Site1Mut
Site2Mut
Site1and2Mut
B
Control miRNA
miR132
miR212
Untreated
Relative expression
TMEM106B
HMBS
C
Part 3'UTR
Full 3'UTR
Site1 Mutant
Site2 Mutant
Site1+2 Mutant
TMEM106B
GAPDH
Protein expression
ns
Part3'UTR
Full3'UTR
Site1Mut
Site2Mut
Site1+2Mut
Vector
D
RNA expression (norm to 2 housekeepers)
E
Site1Mut
Site1
Vector
Luminescence ratio
Control miRNA
miRNA-132
miRNA-212
None
F
Site2Mut
Site2
Vector
miR-132
miR-212
G
Fos
pre-miR-132
pre-miR-212
Expression (+BDNF/-BDNF)
0hr
1hr
2hr
24hr
48hr
72hr
Time after Stimulation
H
miR132
TMEM106B
p250GAP

Figure 2.4. *TMEM106B* is regulated by miR-132 and miR-212.
(A) MiR-132 and miR-212 are predicted to target the same mRNAs through a common seed region (nucleotides in red). The *TMEM106B* 3'UTR has two 8mer miR-132/miR-212 target sites at positions 5084 and 5662 (NM_001134232.1). Bolded, underlined nucleotides indicate regions of perfect complementarity between the *TMEM106B* 3'UTR and miR-132, which may have greater affinity than miR-212 for Site 2 of the *TMEM106B* 3'UTR through a stretch of complementarity outside the seed region. Constructs containing *TMEM106B* with the intact 3'UTR (Full 3'UTR), or truncations or targeted deletions removing one or both miR-132/212 sites, were used for experiments shown in C and D.
(B) Endogenous *TMEM106B* mRNA transcript is significantly reduced with addition of miR-132 or miR-212 mimics (means±SEM from 4 independent transfections).
(C) Compared with the truncated construct (Part 3'UTR), protein levels of the *TMEM106B* construct containing both miR-132/212 binding sites (Full 3'UTR) was decreased by 65%. Deletion of miR-132/212 Site 1 restored TMEM106B protein levels by ~40%, while deletion of miR-132/212 Site 2 had minimal effect. Targeted deletions of both miR-132/212 binding sites (Site1and2Mut) resulted in even greater restoration of TMEM106B expression. Representative immunoblot (top) and quantitation (means±SEM of 5 independent replicates) are shown. For TMEM106B constructs, only the 75kD band was apparent for cell lysates, which were always kept on ice.
(D) Corresponding *TMEM106B* mRNA levels (means±SEM of 5 independent replicates) for the same constructs.
(E,F) Luciferase reporters were generated containing *TMEM106B* 3'UTR miR-132/212 binding Site 1 (Panel E) or Site 2 (Panel F) with 5-10 flanking base pairs. Transfection of MiR-132 or miR-212 resulted in significant decreases in reporter activity (means±SEM from 3 independent replicates) when either site was intact. No change was seen for reporters containing scrambled versions of the miR-132/miR-212 binding sites (Site1Mut and Site2Mut) or for reporters without miR-132 binding sites (Vector). Luminescence ratio is the ratio between the firefly luciferase reporter under 3'UTR control and the constitutively active renilla luciferase reporter, controlling for differences in transfection efficiency.
(G,H) BDNF was applied to neuronally-differentiated SHSY5Y cells to induce CREB-responsive genes. Pre-miR-132 and pre-miR-212, as well as the canonical CREB-responsive gene *fos*, were induced by BDNF within 1-2 hours (Panel G). Mature miR-132 was maximally induced 24 hours after BDNF treatment (Panel H). Expression of the known miR-132 target gene *p250GAP*, as well as *TMEM106B*, decreased with BDNF treatment, with maximal repression at 48 hours for *TMEM106B*. MiR-132 levels had a significant effect on both *TMEM106B* ($p=0.009$) and *p250GAP* ($p=0.006$) expression.
Ratios of BDNF-treated to non-treated conditions are shown for a minimum of 4 separate transfections.
All panels: *$p<0.05$, **$p<0.01$, ***$p<0.001$

132 and miR-212, consistent with CREB-activated transcription of these two miRNAs (**Figure 2.4G**).

We observed a subsequent rise in mature forms of miR-132 and miR-212, peaking at 24 hours after BDNF treatment. Concomitant with the rise in miR-132, *TMEM106B* transcript levels fell, with a maximum effect 48 hours after BDNF treatment (**Figure 2.4H**). Reduction in *TMEM106B* expression was significantly associated with rise in miR-132, and similar in magnitude to observed reduction in the known miR-132 target gene *p250GAP* (Remenyi et al., 2010; Wayman et al., 2008),

consistent with miR-132 repression of *TMEM106B* in a neuronal setting (**Figure 2.4H**). Similar trends were seen for miR-212, but total levels were 5-10 times lower than miR-132 in SHSY5Y cells and negative correlation with *TMEM106B* and *p250GAP* expression was not significant for miR-212 (data not shown). Taken together, these results provide empirical confirmation of the computationally predicted regulation of *TMEM106B* by miR-132 and miR-212.

2.3.4 Neuronal TMEM106B is localized to late endosomes or lysosomes

Having established that the miR-132 cluster is decreased in FTLD-TDP, that TMEM106B is increased in FTLD-TDP, and that miR-132 and miR-212 can both repress *TMEM106B* expression, we turned our attention to the biology of this previously uncharacterized protein. Using immunofluorescence microscopy, we first evaluated the subcellular localization of TMEM106B using the N2077 antibody. Of note, the commercially available antibodies against TMEM106B did not show co-localization with antibodies against various tags upon expression of tagged constructs, so we did not use them for cell biological experiments. FLAG-tagged TMEM106B in HEK293 cells (**Figure 2.5A**) co-localized with the late endosomal/lysosomal marker LAMP-1, supporting previous reports using immortalized cell lines that TMEM106B may be localized to lysosomes (Lang et al., 2012). We extended these findings by demonstrating that endogenous TMEM106B in non-transfected murine primary cortical neurons (**Figure 2.5B**) also co-localized with LAMP-1, as well as Lysotracker, an indicator of acidic organelles (**Figure 2.5C**). In contrast, there was little overlap in staining between TMEM106B and Golgi markers (**Figure 2.5D**) or between TMEM106B and TDP-43 (**Figure 2.5E**). Similar late endosomal/lysosomal staining patterns were seen for murine primary hippocampal neurons, and for primary hippocampal and primary cortical neurons from rats (data not shown). Thus, neuronal TMEM106B is localized primarily in late endosomes or lysosomes.

2.3.5 Over-expression of TMEM106B results in abnormalities in the endosomal-lysosomal pathway

Having established the late endosomal/lysosomal subcellular localization of TMEM106B, we next evaluated the effect of over-expression of TMEM106B on the endosomal-lysosomal

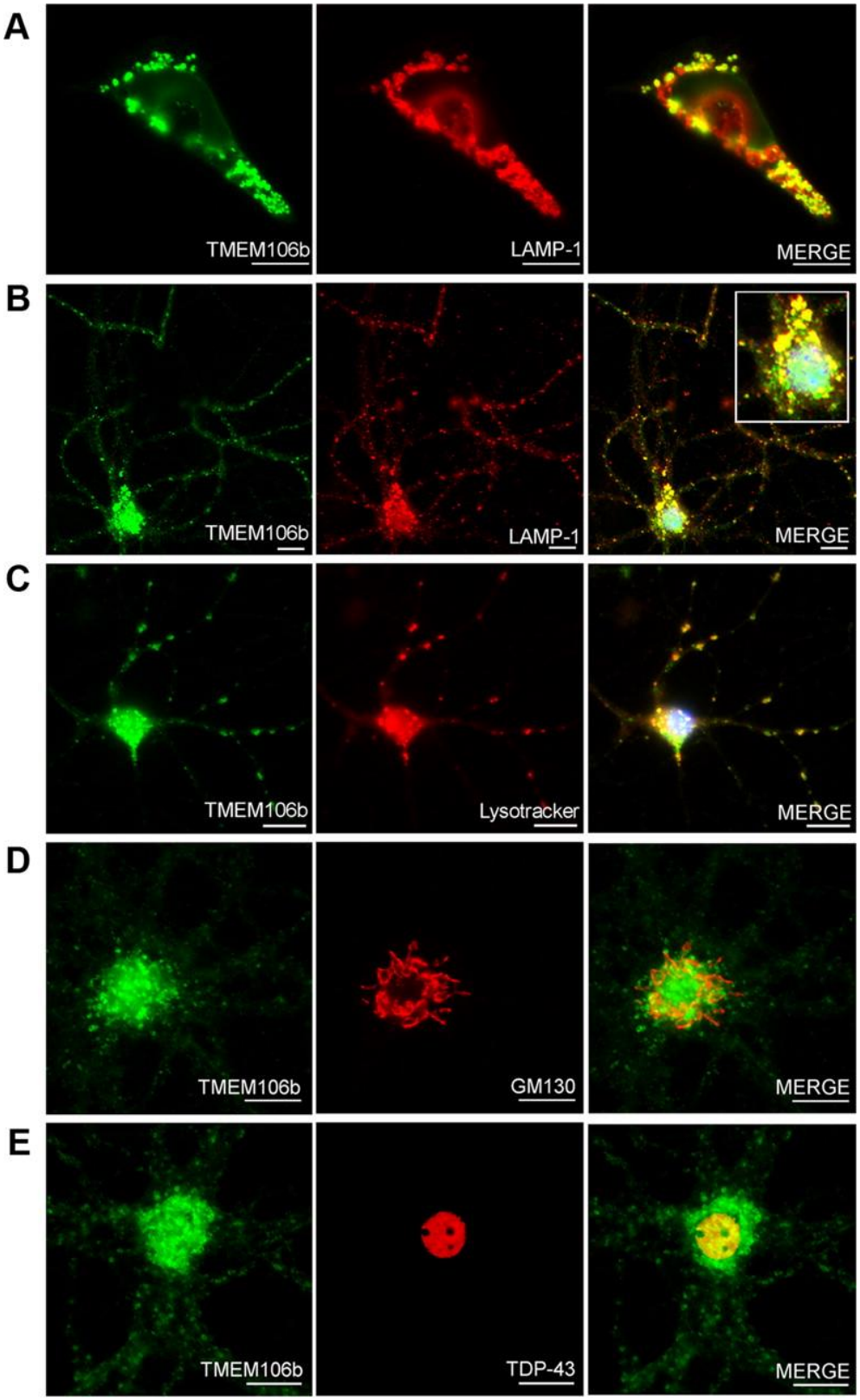

A
TMEM106b
LAMP-1
MERGE
B
TMEM106b
LAMP-1
MERGE
C
TMEM106b
Lysotracker
MERGE
D
TMEM106b
GM130
MERGE
E
TMEM106b
TDP-43
MERGE

Figure 2.5. TMEM106B is associated with late endosomes/lysosomes in neurons.
TMEM106b antibody N2077 was used for immunofluorescence microscopy. Scale bars represent 10µm.
(A) Structures staining with TMEM106B antibody (green) also expressed the late endosomal/lysosomal marker LAMP-1 (red) in HEK293 cells transiently transfected with TMEM106B.
(B) In non-transfected primary murine cortical neurons, endogenous TMEM106B (green) co-localized with LAMP-1 (red) in cell processes, and in the cell body (inset).
(C) In non-transfected primary murine cortical neurons, TMEM106B (green) co-localized with a marker for acidic organelles, Lysotracker (red), corroborating the association of TMEM106B with late endosomes/lysosomes.
(D) TMEM106B did not co-localize with the cis-Golgi marker GM130.
(E) TMEM106B did not co-localize with TDP-43.

pathway. To do this, we transiently over-expressed TMEM106B and performed double-label immunofluorescence for TMEM106B and LAMP-1. As shown in **Figure 2.6A**, HeLA cells over-expressing TMEM106B (arrowhead) demonstrated enlargement of LAMP-1+/TMEM106B+ organelles, while neighboring cells with normal levels of TMEM106B did not. We observed this general enlargement of LAMP-1+ organelles upon TMEM106B over-expression in multiple cell lines, including HEK293 cells, HeLA cells, and the mouse neuronal cell line Neuro2A, with virtually no "normal-sized" LAMP-1+ organelles (**Figure 2.6B**) seen. Additionally, we occasionally observed the formation of very large cytoplasmic vacuolar structures ~5µm in diameter; these structures showed LAMP-1 and TMEM106B staining co-localized at the limiting membrane (**Figure 2.6A**, bottom).

Since large vacuolar structures have been described upon treatment of cells with alkalizing agents such as the lysosomotropic drug chloroquine (Brown et al., 1984), we sought to determine whether these large LAMP-1+/TMEM106B+ organelles were acidified. To do this, we used the pH-sensitive dyes Lysotracker and Lysosensor, both weak bases that fluoresce intensely at low pH and weakly (or not at all) at higher pH. As shown in **Figure 2.6C**, HeLA cells over-expressing TMEM106B (arrowhead) contained large LAMP-1+/TMEM106B+ organelles which were not as acidic as the LAMP-1+ organelles in neighboring cells with normal levels of TMEM106B expression. Indeed, the Lysotracker mean fluorescence intensity of the cytoplasmic compartment of cells over-expressing TMEM106B was significantly lower than that of cells expressing normal levels of TMEM106B, indicating a general failure to acidify organelles (**Figure 2.6D**). The effect of

TMEM106B over-expression was reminiscent of the cellular alkalization induced by bafilomycin A1 (BafA1), a selective inhibitor of the vacuolar ATPase that acidifies lysosomes (**Figures 2.6C** (bottom) and **2.6D**). Similar results were obtained in HEK293 cells, and in both HeLAs and HEK293 cells live-imaged with the Lysosensor dye (data not shown).

Alkalization of cells with BafA1 was recently reported to result in increased TMEM106B expression (Lang et al., 2012). We confirmed this finding in HeLA cells (**Figure 2.6E**) and HEK293 cells (data not shown). Thus, not only can TMEM106B over-expression impair endosomal-lysosomal acidification, but impairing the acidification of these organelles can in turn result in increased expression of TMEM106B.

We sought to determine whether these enlarged, poorly acidified LAMP-1+ organelles were improperly acidified lysosomes, or late endosomes unable to mature into lysosomes. To do so, we performed double-label immunofluorescence for LAMP-1 and the cation-independent mannose-6-phosphate-receptor (M6PR), which is not present on lysosomes, under conditions of TMEM106B over-expression. In HeLA cells at steady state, the M6PR was found primarily in the trans-Golgi network (TGN) and did not colocalize extensively with LAMP-1 (**Figure 2.6F**, top). However, in cells over-expressing TMEM106B, LAMP-1 and the M6PR co-localized extensively in enlarged organelles (**Figure 2.6F**, bottom), suggesting that there may be a block in endosomal-lysosomal fusion, and, further, that retrograde transport from the late endosome to the TGN is affected. Taken together, our data suggest that over-expression of TMEM106B results in endosomal-lysosomal dysfunction as demonstrated by dramatic abnormalities in late endosome-lysosome morphology and acidification, and M6PR trafficking. Moreover, blocking endosomal-lysosomal function could itself prevent TMEM106B degradation, causing its levels to increase, further exacerbating these defects.

2.3.6 Over-expression of TMEM106B alters the appearance and compartmentalization of progranulin

Since progranulin has been previously linked to lysosomes through proteomic identification of lysosomal contents (Kollmann et al., 2005), we investigated whether TMEM106B and progranulin

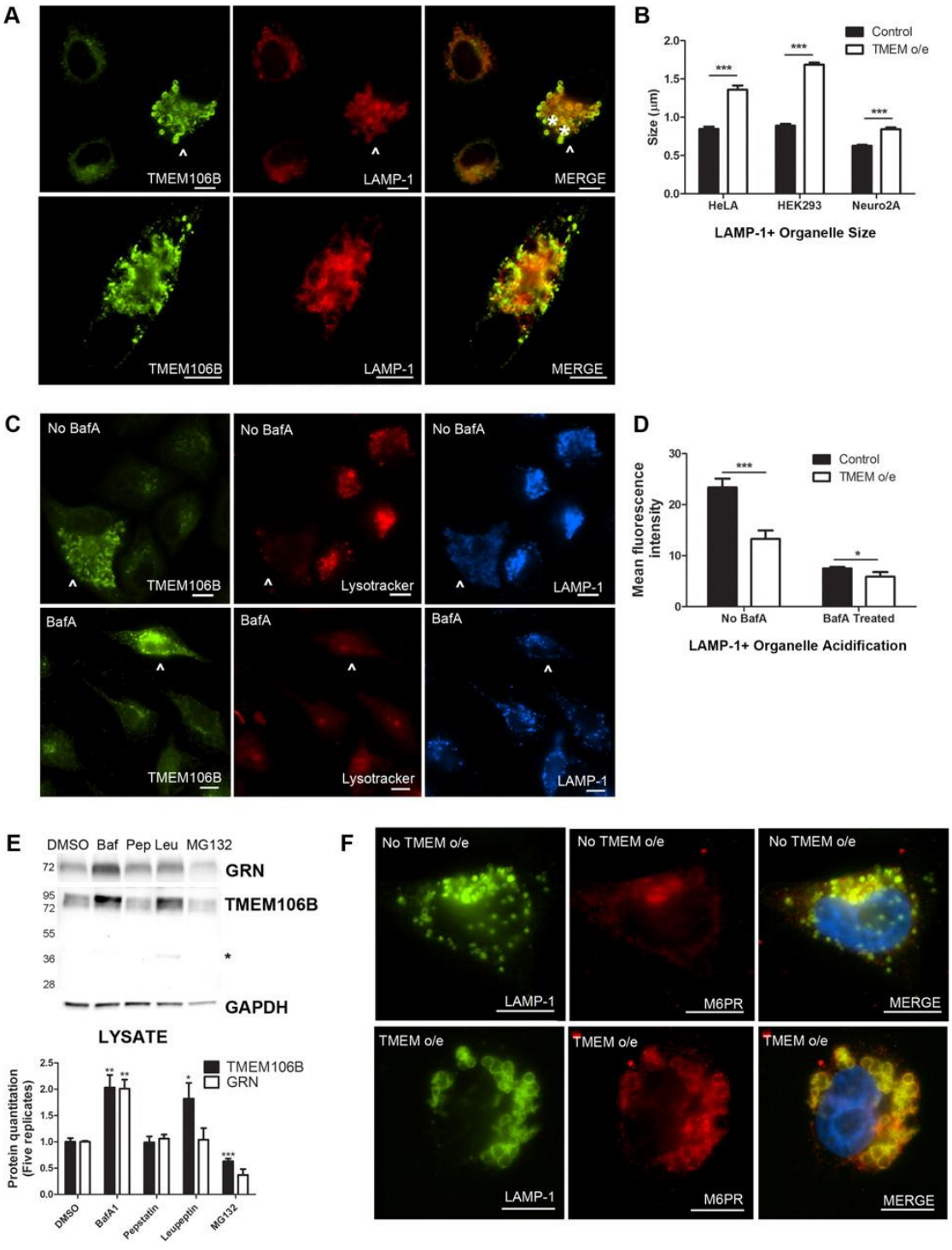


Figure 2.6. Over-expression of TMEM106B results in abnormalities in the endosomal-lysosomal pathway.

(A) In HeLa cells over-expressing TMEM106B (arrowhead), LAMP-1+ organelles demonstrate a general increase in size, compared with neighboring cells not over-expressing TMEM106B. In addition, over-expression of TMEM106B also results in occasional formation of large vacuolar

structures ~5µm in diameter (asterisks indicate two vacuolar structures in top panel, also pictured throughout the cytoplasm of cell in bottom panel). While these large vacuolar structures occur only occasionally with TMEM106B over-expression (the more typical finding is enlarged ~1.5µm LAMP-1+ organelles), they are not seen in the absence of TMEM106B over-expression.
(B) Similar results were obtained in HeLas, HEK293 cells, and in the neuronal cell line Neuro2A. Size quantitation (means +/- SEM) was performed by measuring LAMP-1+ organelle diameter on >10 40X fields containing a mixture of cells with and without TMEM106B over-expression. Because the large vacuolar structures are only occasionally seen, they were not included in the quantitation.
(C) HeLa cells over-expressing TMEM106B (arrowhead) showed less intense staining with Lysotracker, a dye which demonstrates greater fluorescence at lower pH, than neighboring cells not over-expressing TMEM106B (top). This effect was abrogated by treatment of cells with bafilomycin A1, an inhibitor of the vacuolar ATPase, which resulted in diminished Lysotracker fluorescence for all cells (bottom).
(D) Quantitation of mean fluorescence intensity for cells over-expressing TMEM106B demonstrated that Lysotracker staining was significantly less intense than in neighboring cells with normal levels of TMEM106B expression. Quantitation (means +/- SEM) was performed on >10 40X fields containing a mixture of cells with and without TMEM106B over-expression.
(E) Immunoblot analysis of HeLa cells treated with the vacuolar ATPase inhibitor bafilomycin A1 (Baf) showed increased intracellular levels of TMEM106B and progranulin. Treatment with the lysosomal protease inhibitor leupeptin (Leu) increased levels of TMEM106B but did not affect levels of progranulin. Treatment with the lysosomal protease inhibitor pepstatin A (Pep) did not affect either protein, while treatment with the proteasome inhibitor MG132 decreased TMEM106B levels. Representative immunoblot (top) and quantitation of five replicate immunoblots (means +/- SEM, bottom) are shown. Asterisk indicates TMEM106B 40kD band only seen with leupeptin treatment.
(F) Under normal conditions, the cation-independent mannose-6-phosphate receptor (M6PR) does not co-localize with LAMP-1. In cells over-expressing TMEM106B, M6PR co-localizes with LAMP-1 at the limiting membrane of enlarged LAMP-1+ organelles.
*p<0.05, **p<0.01, ***p<0.001.
All immunofluorescence panels: TMEM106B staining performed with N2077. Scale bar represents 10µm.

are found together in lysosomes. Indeed, double label immunofluorescence showed that TMEM106B partially colocalized with progranulin in murine primary cortical neurons (**Figure 2.7A**). Moreover, endogenous neuronal TMEM106B and progranulin appeared most co-localized in neuronal processes, within LAMP-1+ organelles (**Figure 2.7B**).

Defects in the production or secretion of progranulin are known to cause FTLD-TDP and among the cases with progranulin mutations, we observed increased mRNA and protein expression of TMEM106B (**Figure 2.2**). Furthermore, progranulin has been reported to internalize via endocytosis to lysosomes using sortilin-1 as a plasma membrane receptor (Hu et al., 2010), and our data suggested that TMEM106B over-expression may impair the endosomal-lysosomal pathway. Finally, impairment of endosomal-lysosomal acidification, shown here to be an effect of TMEM106B over-expression, has been reported by others (Capell et al., 2011) to result in

increased progranulin levels, a finding we confirmed (**Figure 2.6E**). Thus, we asked what impact increased TMEM106B expression might have on progranulin.

Intriguingly, over-expression of TMEM106B in HEK293 cells (**Figure 2.7C**, arrowhead) and HeLA cells resulted in striking changes in the appearance of progranulin (**Figure 2.7C**), in addition to the enlargement of LAMP-1+ organelles. Indeed, ~60% of cells over-expressing TMEM106B contained intense puncta that colocalized with progranulin, while less than 5% of cells with normal levels of TMEM106B expression showed this pattern (**Figure 2.7D**). In neighboring, non-transfected cells with normal levels of TMEM106B expression (**Figure 2.7C**, arrows), progranulin intracellular staining was much less intense, exhibiting the more diffuse pattern previously reported in the literature (Shankaran et al., 2008).

These observations suggested that increased expression of TMEM106B might alter the trafficking of progranulin. We therefore tested the hypothesis that TMEM106B over-expression would affect levels of intracellular or extracellular progranulin. As shown in **Figure 2.7E**, intracellular progranulin levels significantly increased by ~30% with over-expression of TMEM106B in HEK293 cells. In contrast, extracellular progranulin measured in the conditioned medium remained relatively constant, showing a non-significant slight decrease in the setting of TMEM106B over-expression. Normalization of progranulin measurements to those of another secreted protein, IGFBP-2, did not alter these results and in fact increased the apparent differences. Thus, over-expression of TMEM106B increases levels of intracellular progranulin.

Taken together, these data demonstrate that aberrant over-expression of TMEM106B affects the distribution and intracellular levels of progranulin, suggesting that the two proteins may act in the same pathogenic pathway in FTLD-TDP.

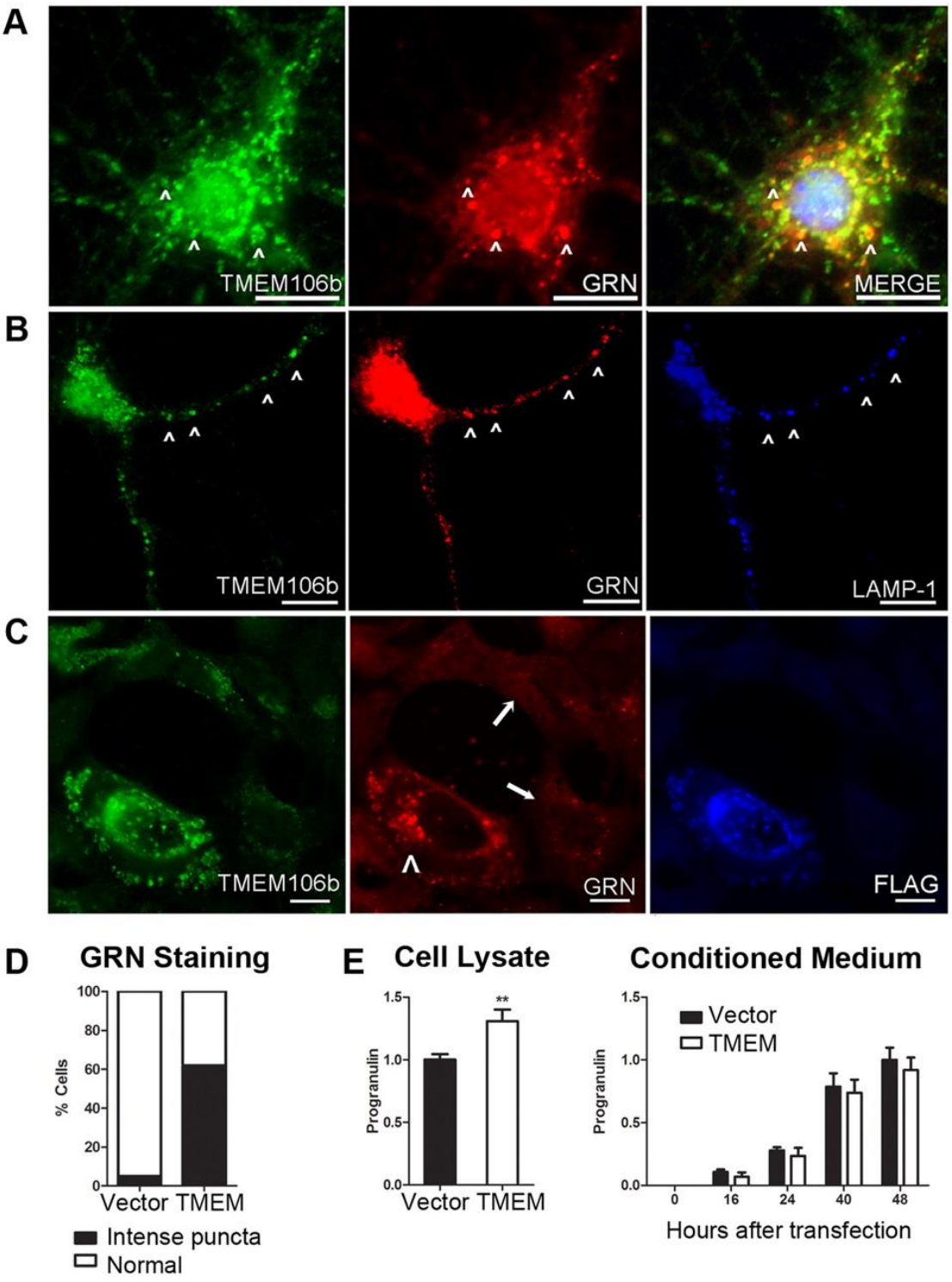


Figure 2.7. Over-expression of TMEM106B alters the compartmentalization of progranulin. Immunofluorescence microscopy (panels A-C) performed on cells stained for TMEM106B (N2077 antibody) and progranulin. Scale bar represents 10µm.

(A,B) Endogenous TMEM106B (green) in non-transfected primary murine cortical neurons co-localized with progranulin (GRN, red) in the cell body (A, arrowheads), and in processes (B, arrowheads). TMEM106B and GRN co-localized within late endosomes or lysosomes, as indicated by LAMP-1 staining (blue, panel B).
(C) Progranulin (GRN, red) appearance changed under conditions of TMEM106B over-expression. Progranulin formed intensely stained cytoplasmic puncta variably co-localizing with TMEM106B (green) only in HEK293 cells over-expressing FLAG-tagged TMEM106B (arrowhead). In the absence of TMEM106B over-expression, progranulin staining was much less intense (arrows).
(D) More than 60% of cells over-expressing TMEM106B showed intense cytoplasmic puncta of progranulin, compared with <5% of cells with normal levels of TMEM106B expression. Assessment of progranulin staining pattern performed on six 20X fields of HEK293 cells containing a mixture of cells with and without TMEM106B over-expression.
(E) Intracellular (left) and extracellular/secreted (right) pools of progranulin were measured by ELISA under conditions of TMEM106B over-expression (TMEM, white bars) vs. vector transfection in HEK293 cells. Progranulin measurements (means ±SEM for 5 experiments) were normalized to total protein in the cell lysate, to account for differential rates of cell growth. Over-expression of TMEM106B resulted in a 30% increase in intracellular progranulin, with a trend towards decreased extracellular progranulin. Intracellular progranulin is shown measured at 48 hours after transfection of TMEM106B. Extracellular progranulin is shown at baseline and indicated time periods after transfection of TMEM106B. **$p<0.01$

2.4 DISCUSSION

Following our recent GWAS reporting the uncharacterized gene *TMEM106B* as a risk factor for FTLD-TDP, here we describe *TMEM106B*'s genetic regulation and relationship to progranulin. Specifically, we show that TMEM106B is highly expressed in human brain tissue, with particularly increased expression in *GRN*(+) FTLD-TDP. We further demonstrate that FTLD-TDP (with and without *GRN* mutations) shows <50% normal expression of three members of the miR-132 cluster: miR-132, miR-132*, and miR-212. Strikingly, miR-132 and miR-212 repress *TMEM106B* through two specific sites in the *TMEM106B* 3'UTR. Having observed increased *TMEM106B* expression in FTLD-TDP, and uncovered a possible upstream cause in miR-132/212, we then explored the consequences of *TMEM106B* over-expression. Over-expression of TMEM106B, which shows steady-state neuronal localization to late endo-lysosomes and co-localizes in these organelles with progranulin, disrupts late endo-lysosome function and consequently, both the appearance and compartmentalization of progranulin. Taken together, the data presented in the current study are compatible with a model (**Figure 2.8**) which illustrates how decreased levels of miR-132/212 lead to increased *TMEM106B* expression, perturbation of progranulin pathways, and increased risk of developing FTLD TDP.

There have been conflicting data regarding the levels of TMEM106B expression in FTLD-TDP. In our initial paper (Van Deerlin et al., 2010), we reported that FTLD-TDP patients show increased expression of *TMEM106B* mRNA in frontal cortex. Others, however, have not found significant differences in *TMEM106B* mRNA expression in FTLD-TDP patients vs. controls (van der Zee et al., 2011). In the present study, we demonstrate again that TMEM106B expression is increased in disease states. First, we directly measure human brain TMEM106B at the mRNA and protein levels, quantifying transcript levels in multiple brain regions; we consistently observe higher levels of TMEM106B in *GRN*(+) FTLD-TDP. In addition, we present a second, independent line of evidence suggesting that TMEM106B expression is aberrantly elevated in FTLD-TDP. Specifically, we find that miR-132 and miR-212, both computationally predicted and experimentally shown to negatively regulate *TMEM106B*, are significantly reduced in *GRN*(-) and *GRN*(+) FTLD-TDP brain.

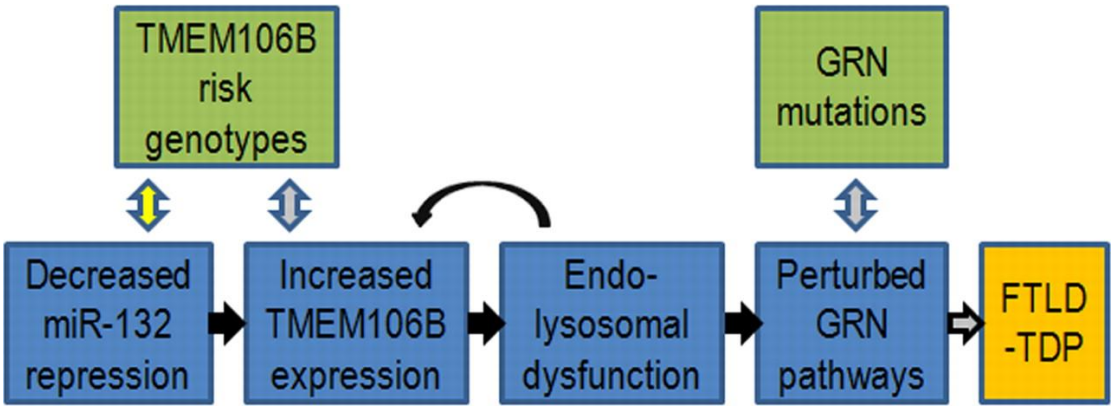


Figure 2.8. Hypothetical model of causes and effects of TMEM106B over-expression in FTLD-TDP.
Our data are compatible with a model whereby decreased levels of miR-132/212 result in increased TMEM106B expression. As a result, increased TMEM106B expression leads to (1) endosomal-lysosomal dysfunction, which may in turn further increase levels of TMEM106B, and also to (2) perturbation of progranulin pathways, thereby increasing the risk of developing FTLD-TDP. Black arrows indicate steps evidenced by the current study, grey arrows indicate steps reported in the literature, and yellow arrows indicate hypothetical steps.

The expected effect of a reduction in these miRNAs would be de-repression, or over-expression, of their target mRNA *TMEM106B*.

In terms of miR-132/212 regulation of TMEM106B, while both miRNAs can decrease TMEM106B expression, miR-132 may be the dominant miRNA. Specifically, miR-132 levels are orders of magnitude higher than miR-212 levels in human brain tissue, and only miR-132 elevation was significantly associated with decrease of *TMEM106B* in neuronal cells.

The miR-132 cluster – comprised of miR-132, miR-132*, miR-212, and a fourth potential member, miR-212* – has been previously shown to be important in neuronal development and in the neurobiology of addictive behavior. Much of the literature has focused on miR-132 and miR-212, which are also the only two miRNAs in the cluster that show potential binding to *TMEM106B*. Specifically, introduction of miR-132 in primary neuronal cultures has been reported to stimulate neurite outgrowth through specific effector mRNAs (Vo et al., 2005), and knockdown of the miR-132 cluster in newborn hippocampal neurons decreases the length and arborization of dendrites (Magill et al., 2010). Increasing striatal expression of miR-212 through lentiviral delivery, on the other hand, has been reported to decrease addictive cocaine-ingestion behavior in rats, while

inhibition of miR-212 has an opposing effect (Hollander et al., 2010). Both miR-132 and miR-212 appear to be regulated by the transcription factor CREB (Impey et al., 2004), but the connection, if any, between the neuronal phenotypes observed with miR-132 manipulation and the whole organism behaviors observed with miR-212 manipulation are unclear. In any case, miR-132 and miR-212 have been postulated to be the downstream effectors of the phenotypes – including increased neuronal survival – associated with CREB activation (Magill et al., 2010; Wayman et al., 2008). Thus, the greatly decreased levels of miR-132 cluster expression observed in FTLD-TDP may have deleterious effects on neurons beyond the de-repression of *TMEM106B*.

In addition to demonstrating the miR-132/212 regulation of TMEM106B, our study adds to a growing body of evidence that TMEM106B and progranulin may be mechanistically related. That is, *TMEM106B* genotype has been reported to correlate with plasma levels of progranulin in normal controls (Finch et al., 2011), and with age at onset in *GRN*(+) FTLD-TDP (Cruchaga et al., 2011). Evidence from the current study to support a link between TMEM106B and progranulin includes: 1) increased protein and mRNA expression of TMEM106B in *GRN*(+) FTLD-TDP cases; 2) abnormal localization of TMEM106B in neuronal processes in *GRN*(+) FTLD-TDP cases; 3) co-localization of TMEM106B and progranulin in late endo-lysosomes; and 4) increased accumulation of intracellular progranulin when TMEM106B is overexpressed. Indeed, we provide the first demonstration of a relationship between TMEM106B and progranulin in a model system which can be directly manipulated. Of note, a prior study (Lang et al., 2012) reported no change in progranulin levels with over-expression of TMEM106B, a result that contradicts our present result. Various methodological differences, such as our use of ELISA to measure progranulin, may explain the discrepancy.

At a more general level, the data presented here strongly support *TMEM106B* as a risk gene for FTLD-TDP and the source of the 7p21 signal found by GWAS. First, there is the striking fact that both miR-132 and *TMEM106B* emerged as single-best hits from independent, genomic-scale screens in FTLD-TDP. Their convergence as a microRNA:mRNA regulatory pair greatly adds to our confidence that both are real signals and also argues for the centrality of the miR-132/212:

TMEM106B pathway in this disease. Second, we link TMEM106B to effects on progranulin – haploinsufficiency of which is a known pathogenic mechanism in FTLD-TDP. In doing so, our data suggest a clear direction for further mechanistic studies of TMEM106B. It is worth noting that establishment of whether a GWAS signal is "true" and, additionally, what the causative gene behind a signal might be, is not always straightforward. Indeed, while additional cohorts of clinical FTLD (van der Zee et al., 2011) and clinical-pathologic FTLD/FTLD-TDP cases (Finch et al., 2011) have replicated the association of FTLD with rs1990622 genotype, some clinical cohorts have not demonstrated this association (Rollinson et al., 2011a), possibly because only ~50% of FTLD clinical cases have underlying FTLD-TDP.

Future directions suggested by the data presented here include a detailed analysis of the biochemistry of TMEM106B. Specifically, we have shown that TMEM106B is glycosylated and heat-sensitive, with both 75kD and 40kD species -- collapsing to ~60kD and 31kD, respectively, after deglycosylation -- apparent by immunoblotting. However, it is presently unclear whether TMEM106B, with a predicted molecular weight of 31kD, migrates at the higher molecular weight because it self-associates into a dimer or because it is complexed to another unknown protein. Understanding the native complex in which TMEM106B occurs may be important in determining its molecular function in normal and disease states.

Additionally, the exact molecular mechanisms by which increased expression of TMEM106B affects progranulin are at present unclear and would be an important direction for future studies. Several hypotheses are compatible with our data. First, increased TMEM106B might cause intracellular retention of progranulin, an intriguing possibility since decreased secretion of progranulin is a known mechanism of FTLD-TDP pathogenesis (Mukherjee et al., 2008; Shankaran et al., 2008). We note here that while we did not see a significant drop in secreted progranulin with TMEM106B over-expression, the trend towards decreased extracellular progranulin was present at every timepoint, and technical limitations such as the high turnover rate in culture medium may have precluded our ability to see a significant change in levels of secreted progranulin. Second, increased TMEM106B might increase the internalization of progranulin from

the extracellular space, in the same way that sortilin-1 over-expression increases progranulin's internalization and sorting into lysosomes (Hu et al., 2010). Third, increased TMEM106B may impair the maturation of lysosomes, as suggested by our observation of enlarged, poorly acidified organelles expressing both LAMP-1 and the M6PR upon TMEM106B over-expression. This defect may in turn alter the dynamics of progranulin internalization and turnover, as others have shown (Capell et al., 2011), and we have confirmed, that blocking lysosomal acidification can increase levels of progranulin.

In conclusion, we demonstrate here the microRNA regulation of TMEM106B and its effects on the endosomal-lysosomal pathway and the FTLD-TDP-associated protein progranulin. Our findings thus open up novel directions for elucidating miRNA-based mechanisms of neurodegeneration in FTLD-TDP and related TDP-43 proteinopathies that could become targets for drug discovery in these currently untreatable diseases.

2.5 METHODS

Human brain samples

Frontal cortex samples from 12 FTLD-TDP cases (5 with *GRN* mutations and 7 without *GRN* mutations) and 6 neurologically normal controls of either sex (see **Table 2.1** for details) were obtained from the University of Pennsylvania Center for Neurodegenerative Disease Research Brain Bank. Total RNA was isolated and evaluated for quality control parameters as previously described (Chen-Plotkin et al., 2008), with the exception that a column purification step was not used, in order to retain small RNAs. Protein was sequentially extracted from a subset of frontal cortex samples. Informed consent was obtained for postmortem studies.

Of note, some of the frontal cortex samples used for mRNA quantitation were previously reported in our GWAS study (Van Deerlin et al., 2010); these data were included here so that sets of data from multiple brain regions included the same samples.

MicroRNA screening and QRT-PCR validation.

1µg of total RNA from each individual brain sample, as well as 1µg of a pooled reference sample, was hybridized to the miRCURY LNA array version 11.0 (Exiqon, Copenhagen) for microRNA quantitation. No microRNA enrichment was needed, as concentrations of miRNAs were high. Statistical analyses of miRNA expression were performed using open source R software packages available from Bioconductor and specifically the limma package for two-color arrays. Microarray QC was performed as previously described (Chen-Plotkin et al., 2008); no outlier chips were identified for elimination. Raw data were RMA normalized (Wettenhall and Smyth, 2004), and median values for each microRNA were used to compare groups using pairwise contrasts within an ANOVA (Analysis of Variance) model correcting for gender and age. R-scripts for these analyses are available on request. Promising candidate microRNAs found by array screening to differ in disease were evaluated using QRT-PCR with TaqMan microRNA assays from Applied Biosystems (AB Assay ID 000457, AB Assay ID 002132, AB Assay ID 000515); microRNAs were normalized to either the geometric mean of two housekeeping small RNAs (human let-7a, AB

Assay ID 000377; human RNU38b, AB Assay ID 001004) or a neuron-specific microRNA (human miR-124a, AB Assay ID 000446).

QRT-PCR was also used for transcript quantitation for *TMEM106B* and other mRNAs of interest as previously described (Chen-Plotkin et al., 2008). Relative quantification was performed (delta-delta method) using ß-actin and/or PPIA as our reference housekeeping gene after verification that these transcripts remain constant in the experimental conditions used.

SNP genotyping

DNA was extracted from brain samples and genotyping was performed using TaqMan chemistry-based allelic discrimination assays as previously described (Van Deerlin et al., 2010).

MicroRNA target prediction and verification

The TargetScan program (release 5.1, http://www.targetscan.org) was used to predict targets of specific microRNAs and specific microRNA regulators of TMEM106B. To verify the predicted regulation of TMEM106B by miR-132 or miR-212, we performed three types of assays.

First, in an endogenous TMEM106B assay, HEK293 cells were plated one day before transfection at 2 x 10^5 per well in 12-well plates. The following day, 75pmoles of miRNA mimic (miR-132, Applied Biosystems PM10166; miR-212, Applied Biosystems PM10340; miR-124 Applied Biosystems PM 10691) were transfected using Lipofectamine 2000 (Invitrogen) according to manufacturer's instructions in serum-free conditions. Medium was replaced after 4-6 hours to standard 10% FBS-containing medium. Appropriate amounts of microRNA mimics were determined empirically by testing a range of concentrations. 48 hours after transfection, cells were harvested, and total RNA was isolated using the miRNeasy kit (Qiagen) before quantitation by QRT-PCR.

Second, *TMEM106B* constructs containing the full 3'UTR, or lacking various portions of the 3'UTR were cloned into a commercially available construct (*TMEM106B* construct, Origene SC113580). For miR-132/212 Site 1 and 2 deletions, the seed match region (red nucleotides on the 3' end of sequences shown in Figure 4A) was deleted. 4µg of the relevant *TMEM106B* construct was transfected into each well of a 6-well plate. RNA and protein were harvested at 48

hours after transfection. For protein studies, cells were lysed in RIPA buffer with protease inhibitors and PMSF on ice, and lysates were centrifuged (*16,000xg*, 4°C, 15min) to remove debris, before downstream immunoblotting.

Third, luciferase reporter constructs were created adding 35-40bp of *TMEM106B* 3'UTR sequence including and flanking the relevant miR-132/212 binding sites to the 3'UTR regulatory region of a commercially available dual-luciferase reporter (pmiRGLO Vector, Promega). In addition, mutated versions of these same constructs, containing the same 35-40bp in a scrambled order to abrogate miR-132/212 binding, were also cloned into the luciferase reporters in a similar manner. For miR-132/212 Site 1, the following sequence (flanking regions plus miRNA binding sites, with miRNA binding site in bold) was inserted: CTAGTGAGA**TGGTGGAGTAAAAAGACTGTTA**AACATTGCA. For miR-132/212 Site 1 mutated sequence, the following sequence (flanking regions plus scrambled miRNA binding sites, with scrambled miRNA binding site in bold) was inserted: CTAGTGAGA**TAGGGGTGAATACAATGAGCTT**AACATTGCA. For miR-132/212 Site 2, the following sequence (flanking regions plus miRNA binding sites, with miRNA binding site in bold) was inserted: CTAGTATTA**CTACCATGTAGACTGTTA**TAGTTTGCA. For miR-132/212 Site 2 mutated sequence, the following sequence (flanking regions plus scrambled miRNA binding sites, with scrambled miRNA binding site in bold) was inserted: CTAGTATTA**GTACTACGTACAGGTATT**TAGTTTGCA. These constructs were then co-transfected into HEK293 cells with microRNA mimics as described above. 50ng of the relevant luciferase reporter construct was co-transfected. 24 hours after transfection, lysates were prepared according to instructions from the Dual Luciferase Assay kit (ProMega) and read on a Berthold LB941 TriStar vTI Multimode Reader.

TMEM106B antibodies

An N-terminus peptide corresponding to amino acid residues 4-19 (SLSHLPLHSSKEDAYDC) was synthesized and used to immunize rabbits, and site-specific IgG was affinity-purified with an N-terminus peptide column. The resulting antibody was named N2077.

Antibody specificity was verified with immunoblots and immunofluorescence microscopy (Figure 1). Peptide preabsorption was performed with a 5:1 (by weight) ratio of immunizing peptide to N2077 antibody, incubated overnight at 4°C, with rotation.

In addition, we tested commercially available antibodies against TMEM106B raised against N-terminus amino acids 1-50 (Bethyl Laboratories, Montgomery, TX) and against the C-terminus of the protein (Proteintech, Chicago, IL). Neither antibody demonstrated specificity in immunocytochemical experiments (data not shown), but the N-terminus antibody recognized the same bands as N2077 on immunoblot (**Figure 2.1D**) and was used to confirm key biochemical findings.

Cell culture and transfection

Primary cortical and hippocampal neurons were prepared from embryonic day 18 (E18) to E20 C57BL/6 mice, plated, and maintained as previously described (Tseng et al., 2006). In addition, HEK293 and SHSY5Y cells were maintained and used for experiments. Where indicated, cells were transfected with the respective construct(s) using Lipofectamine-2000 (Invitrogen) one day after plating (50-70% confluency) according to manufacturer instructions; transfections were performed under serum-free conditions. SHSY5Y cells were neuronally differentiated prior to their use in experiments as previously described (Remenyi et al., 2010). Neuro2A cells were neuronally differentiated directly after transfection for 48 hours prior to their use by incubation in neurobasal media (Gibco catalog # 1208) with B27 (Gibco catalog #1209) together with complete DMEM in a 1:1 ratio in an adaption of a previously described protocol (Ishikura et al., 2005).

Treatment of cells with lysosomal and proteosomal inhibitors

Cells were treated with the vacuolar ATPase inhibitor bafilomycin A1, lysosomal protease inhibitors Pepstatin A or Leupeptin, proteosomal inhibitor MG132, or control DMSO as previously described (Capell et al., 2011), with the exception that Leupeptin was not used in a mix but rather alone at a concentration of 1μM.

Immunofluorescence microscopy

Cells were fixed (2% paraformaldehyde, 15min, RT), washed five times with PBS, and then blocked for 60 minutes at room temperature in blocking buffer (3% BSA, 0.05% saponin, in PBS). Primary antibodies were diluted in blocking buffer, and coverslips were incubated overnight at 4°C. The next day, coverslips were washed four times with blocking buffer before incubation with secondary antibody. Secondary antibodies were diluted in blocking buffer, and coverslips were incubated for 1 hour at room temperature protected from light. Following this incubation, coverslips were washed five times with blocking buffer, once with PBS, and then mounted on slides (ProLong Gold, Invitrogen).

The following antibodies and conditions were used. TMEM106B: N2077 antibody (see "TMEM106B antibodies") was used at 1µg/mL. FLAG: M2 antibody (Sigma) was used at 1 µg/mL. Progranulin: goat anti-human GRN (R&D Systems) was used at 1 µg/mL, and sheep anti-mouse GRN (R&D Systems) was used at 1 µg/mL. TDP-43: N171 antibody (mouse monoclonal antibody generated at University of Pennsylvania (Lippa et al., 2009)) was used at 1 µg/mL. Mannose-6-phosphate receptor (M6PR, cation-independent): mouse anti-human M6PR (Catalog number ab2733, Abcam, Cambridge, MA) was used at 2 µg/mL. LAMP-1: mouse anti-human H4A3 (DSHB, Iowa City, IA) and mouse anti-human CD107A (Catalog number 555798, BD Biosciences, San Jose, CA) were used at 1 µg/mL; rat anti-mouse 1D4B (DSHB, Iowa City, IA) was used at 1 µg/mL. GM130: mouse GM130 antibody recognizing human, mouse, and rat GM130 (BD Biosciences, San Jose, CA) was used at 1.25 µg/mL. All secondary antibodies were Alexa Fluor antibodies (Invitrogen, Carlsbad, CA) used at 1:1000, with the exception of an AMCA anti-mouse IgG (Vector Labs, Burlingame, CA) used at 1:100.

Lysosensor and Lysotracker staining

The weak base dyes Lysosensor DND-189 (Invitrogen, Carslbad, CA) and Lysotracker DND-99 (Invitrogen, Carlsbad, CA) were used to assess the acidity of intracellular organelles, since they both fluoresce intensely at low pH, and more weakly or not at all at higher pH. Live cells were incubated with Lysosensor (1µM, 30 minutes, 37ºC) or with Lysotracker (100nM, 2 hours, 37ºC).

Lysosensor-stained cells were imaged live immediately in 4ºC PBS. Lysotracker-stained cells were imaged live in 4ºC PBS or fixed and stained for immunofluorescence microscopy as above.

Quantification of LAMP-1+ organelle size, Lysosensor/Lysotracker mean fluorescence intensity, and progranulin staining appearance

LAMP-1+ organelle size: An average of twelve fields containing a representative heterogenous mixture of cells with and without TMEM106B over-expression after transient transfection were captured at 40X magnification after staining for LAMP-1 and TMEM106B. The diameter of LAMP-1+ organelles was measured for an average of three TMEM106B over-expressing and three TMEM106B non-over-expressing cells per field, and average diameters (+/- SEM) were calculated for cells with and without TMEM106B over-expression. TMEM106B over-expressing cells occasionally contained very large vacuolar LAMP-1+ organelles; these were not included in the quantitation.

Lysosensor/Lysotracker mean fluorescence intensity: An average of eleven fields containing a representative heterogeneous mixture of cells with and without TMEM106B over-expression after transient transfection were captured at 40X magnification after staining for TMEM106B and the indicated pH-sensitive dye. The cytoplasmic compartment was outlined as a region of interest for measures of mean fluorescence intensity (MFI) in the Lysosensor or Lysotracker channel. An average of two TMEM106B over-expressing and three TMEM106B non-over-expressing cells per field were quantitated for MFI, and average MFIs (+/- SEM) were calculated for cells with and without TMEM106B over-expression.

Progranulin staining appearance: Six fields containing a representative heterogeneous mixture of cells with and without TMEM106B over-expression after transient transfection were imaged at 20X magnification after staining for TMEM106B and progranulin. The presence or absence of intense puncta of progranulin staining was then scored for an average of 14 TMEM106B-over-expressing cells and >50 TMEM106B-non-over-expressing cells per field.

Protein preparation and immunoblotting

Where indicated, sequential biochemical fractionation of cell lysates or postmortem human brain samples was performed. Samples were sequentially extracted in buffers of increasing strength: High-salt (10 mM Tris, 500 mM NaCl, 2 mM EDTA, 1mM DTT, 10% sucrose, pH 7.5), RIPA (50 mM Tris, 150 mM NaCl, 5mM EDTA, 0.5% sodium deoxycholate, 1% NP-40, 0.1% sodium dodecyl sulfate [SDS], pH 8.0), and 2% SDS. Since most TMEM106B is extractable in RIPA with a small fraction in High-salt, we extracted directly into RIPA for routine biochemical experiments unless otherwise indicated. Immunoblotting was performed as previously described (Neumann et al., 2006).

Enzyme-linked immunosorbent assay (ELISA)

Sandwich ELISAs were used for measurement of progranulin and the control secreted protein IGFBP-2 from cell culture lysates and conditioned medium. Specifically, progranulin quantification was performed using a commercially available ELISA (Human progranulin ELISA kit, AdipoGen, Korea). IGFBP-2 quantification was also performed using a commercially available ELISA (Human IGFBP-2 ELISA kit, RayBiotech, Inc, USA).

For the progranulin secretion assay, 50µL samples of conditioned medium were removed from each cell culture well (containing a total of 1mL of medium) at designated time intervals after transfection. At the last time point (48 hours), a 50µL conditioned medium sample was collected, the remaining medium was removed, cells were washed once with PBS, and cell lysates were then extracted in RIPA buffer as described in the preceding section.

Statistical tests

Two-tailed t-tests were used, unless confirming a specific directionality of change (*e.g.* QRT-PCR verification of microarray results), in which case one-tailed t-tests were used. Statistical analysis of microarray data was performed as described in “MicroRNA screening and QRT-PCR validation” section. For BDNF treatment experiments in SHSY5Y cells, linear regressions were used to evaluate the correlation of miRNA levels with target mRNA levels, adjusting for time point of measurements and the interaction of miRNA effect with time. For repeated measurements of

extracellular progranulin over time, repeated measures two-way ANOVA was used to evaluate the effect of increased TMEM106B expression.

CHAPTER 3: TMEM106B IS A GENETIC MODIFIER OF FRONTOTEMPORAL DEMENTIA WITH C9ORF72 HEXANUCLEOTIDE REPEAT EXPANSIONS

by

Michael D. Gallagher, Eunran Suh, Murray Grossman, Lauren Elman, Leo McCluskey, John C. Van Swieten, Safa Al-Sarraj, Manuela Neumann, Ellen Gelpi, Bernardino Ghetti, Jonathan D. Rohrer, Glenda Halliday, Christine Van Broeckhoven, Danielle Seilhean, Pamela J. Shaw, Matthew P. Frosch, International Collaboration for Frontotemporal Lobar Degeneration, John Q. Trojanowski, Virginia M.Y. Lee, Vivianna Van Deerlin, and Alice S. Chen-Plotkin

This chapter was written by Michael Gallagher and Alice Chen-Plotkin and originally published in *Acta Neuropathologica* (March 2014, Volume 127, Issue 3, Pages 407-418). Analyses for genetic modifier effects and progranulin assays were completed by Michael Gallagher and Alice Chen-Plotkin. Genetic data and clinical data were previously collected for another project by the International Collaboration for Frontotemporal Lobar Degeneration.

3.1 ABSTRACT

Hexanucleotide repeat expansions in chromosome 9 open reading frame 72 (*C9orf72)* have recently been linked to frontotemporal lobar degeneration (FTLD) and amyotrophic lateral sclerosis (ALS), and may be the most common genetic cause of both neurodegenerative diseases. Genetic variants at *TMEM106B* influence risk for the most common neuropathological subtype of FTLD, characterized by inclusions of TAR DNA binding protein of 43kDa (FTLD-TDP). Previous reports have shown that *TMEM106B* is a genetic modifier of FTLD-TDP caused by progranulin (*GRN*) mutations, with the major (risk) allele of rs1990622 associating with earlier age at onset of disease. Here we report that rs1990622 genotype affects age at death in a single-site discovery cohort of FTLD patients with *C9orf72* expansions (n=14), with the minor allele correlated with earlier age at death (p=0.024). We replicate this modifier effect in a 30-site international neuropathological cohort of FTLD-TDP patients with *C9orf72* expansions (n=75), again finding that the minor allele associates with earlier age at death (p=0.016), as well as earlier age at onset (p=0.019). In contrast, *TMEM106B* genotype does not affect age at onset or death in 241 FTLD-TDP cases negative for *GRN* mutations or *C9orf72* expansions. Thus, *TMEM106B* is a genetic modifier of FTLD with *C9orf72* expansions. Intriguingly, the genotype that confers decreased risk for developing FTLD-TDP (minor, or C, allele of rs1990622) is associated with earlier age at onset and death in *C9orf72* expansion carriers, providing an example of sign epistasis in human neurodegenerative disease.

3.2 INTRODUCTION

Frontotemporal lobar degeneration (FTLD) is the second most common dementia in individuals under 65 years of age (Seelaar et al., 2011). The most common neuropathological subtype is frontotemporal lobar degeneration with TAR DNA-binding protein of 43kDa (TDP-43) inclusions (FTLD-TDP) (Seelaar et al., 2011). We previously reported the minimally characterized gene, *TMEM106B*, as a risk factor for FTLD-TDP by genome-wide association study (GWAS) (Van Deerlin et al., 2010), and this association has been verified independently (Finch et al., 2011; van der Zee et al., 2011). In our GWAS, three SNPs reached genome-wide significance for association with FTLD-TDP (Van Deerlin et al., 2010); all are located within a 36kb haplotype block that contains *TMEM106B* and no other genes. The major alleles of all three SNPs are associated with increased risk of FTLD-TDP (P=1.08x10^{-11}, odds ratio=1.64 for rs1990622, the top GWAS SNP) (Van Deerlin et al., 2010).

Several studies have begun to elucidate the role *TMEM106B* plays in FTLD-TDP. TMEM106B levels have been shown to be increased in FTLD-TDP brains (Chen-Plotkin et al., 2012; Van Deerlin et al., 2010), and risk-associated alleles resulting in amino acid variation in the TMEM106B protein have been reported to result in higher steady-state levels of TMEM106B through slower protein degradation (Nicholson et al., 2013). In addition, the major allele of rs1990622 has been associated with reduced plasma progranulin (PGRN) levels in both healthy individuals and in individuals with FTLD-TDP caused by mutations in the gene encoding progranulin, *GRN* (Cruchaga et al., 2011; Finch et al., 2011). Mutations in *GRN* are a major cause of familial FTLD-TDP (Gass et al., 2006), and are thought to cause disease via haploinsufficiency of the progranulin protein (Gass et al., 2006; Shankaran et al., 2008). Interestingly, among *GRN* mutation carriers with FTLD (*GRN*+ FTLD), *TMEM106B* risk alleles have been reported to associate with earlier age at disease onset (Cruchaga et al., 2011). Experiments in cell culture systems have also demonstrated that TMEM106B and PGRN co-localize in several cell types, including neurons, and that overexpression of *TMEM106B* alters intra- and extracellular levels of

PGRN (Brady et al., 2013; Chen-Plotkin et al., 2012; Nicholson et al., 2013). Therefore, increased expression of *TMEM106B* may confer risk for FTLD-TDP by altering PGRN levels.

While *GRN* mutations account for ~5% of clinical FTLD cases (Gass et al., 2006), and other rarer, monogenic causes of FTLD are known (including mutations in *MAPT, CHMB2B* and *VCP*) (Hutton et al., 1998; Skibinski et al., 2005; Watts et al., 2004), the majority of familial cases were until recently of unknown cause. This changed in late 2011 when two groups reported that hexanucleotide repeat expansions in the *C9orf72* gene are perhaps the most common cause of familial FTLD, familial amyotrophic lateral sclerosis (ALS), and familial FTLD with motor neuron disease (FTLD-MND) (DeJesus-Hernandez et al., 2011; Renton et al., 2011). Although these mutations display an autosomal dominant mode of inheritance, 3-6% of apparently sporadic cases of FTLD and ALS harbor *C9orf72* expansions as well (DeJesus-Hernandez et al., 2011; Renton et al., 2011).

The function(s) of *C9orf72* and its role in disease are currently areas of ongoing research (Cruts et al., 2013), with evidence for both loss-of-function (Ciura et al., 2013; DeJesus-Hernandez et al., 2011; Gijselinck et al., 2012; Renton et al., 2011) and gain-of-toxic-function (Ash et al., 2013; Fratta et al., 2012; Mori et al., 2013) mechanisms. At a neuropathological level, *C9orf72* expansion positive FTLD (*C9orf72*+ FTLD) and ALS (*C9orf72*+ ALS) cases exhibit TDP-43 pathology reminiscent of *GRN*+ FTLD, as well as mutation-negative ALS and FTLD, although *C9orf72*+ FTLD and ALS cases show unique pathological features as well (Boxer et al., 2011; Snowden et al., 2012; Stewart et al., 2012).

Here, we assess whether *TMEM106B* risk genotypes exert a genetic modifier effect in *C9orf72*+ FTLD and ALS, *GRN*+ FTLD, and FTLD cases without either mutation. We also investigate whether these genotypes are associated with disease status in *C9orf72*+ FTLD and with plasma progranulin levels in *C9orf72*+ expansion carriers.

3.3 RESULTS

3.3.1 *TMEM106B* genotype at rs1990622 influences age at death in a discovery cohort of *C9orf72*+ FTLD

TMEM106B genotype has been shown to demonstrate a genetic modifier effect in FTLD-TDP caused by autosomal dominant mutations in the progranulin gene (*GRN*) (Cruchaga et al., 2011). We therefore asked whether genetic variation at *TMEM106B* influences age at death or age at onset in *C9orf72*+ FTLD or ALS disease cases. We assumed a codominant model for these initial analyses.

In *C9orf72*+ FTLD (n=14), age at death was significantly correlated with *TMEM106B* genotype at rs1990622, the SNP previously found in our GWAS to associate most strongly with FTLD-TDP risk (*P*=0.024, **Table 3.1**). Adjusting for sex and presence/absence of co-existing MND did not affect this association. Moreover, the direction of association was surprising; specifically, the minor allele of rs1990622 (C) was associated with earlier age at death in *C9orf72*+ FTLD. In our prior GWAS, the minor allele of rs1990622 was found to be protective against the development of FTLD.

In contrast, rs1990622 genotype did not affect age at death in *C9orf72*+ ALS (n=39, **Table 3.1**). In this discovery cohort, rs1990622 genotype did not affect age at onset for *C9orf72* expansion carriers who presented with either ALS (n=47) or FTLD (n=26). However, a statistically significant association emerged when we performed a multivariate analysis controlling for gender and presence of FTD in the clinical ALS cases, with the major allele associating with earlier age at onset (n=47, **Table 3.1**).

3.3.2 *TMEM106B* genotype at rs1990622 influences age at onset and age at death in a replication cohort of *C9orf72*+ FTLD

We sought to replicate the genetic modifier effect of *TMEM106B* in *C9orf72*+ FTLD in an independent cohort of patients. Since the majority of cases from our prior FTLD-TDP GWAS had been screened for the presence of *C9orf72* expansions, these cases provided an ideal replication cohort to evaluate the effect of *TMEM106B* rs1990622 genotype on age at death in *C9orf72*+ FTLD

Disease	Outcome	Predictors	Beta (rs1990622, each minor allele)	R^2 for model	*P*-value (rs1990622)
FTLD and FTLD-TDP	Age at Death (n=14)	rs1990622	-6.278	0.303	**0.024 ***
		rs1990622, Sex, MND	-5.297	0.393	**0.049 ***
	Age at Onset (n=26)	rs1990622		n.s.	
		rs1990622, Sex, MND		n.s.	
ALS	Age at Death (n=39)	rs1990622		n.s.	
		rs1990622, Sex, FTD		n.s.	
	Age at Onset (n=47)	rs1990622	4.264	0.044	0.085 n.s.
		rs1990622, Sex, FTD	4.900	0.075	**0.048 ***

Table 3.1. *TMEM106B* genotype affects age at death in *C9orf72* expansion carriers with FTLD or FTLD-TDP in a discovery cohort.
Linear regressions were used to evaluate the effect of *TMEM106B* genotype at rs1990622 on the age at death or age at onset in *C9orf72* expansion carriers from a discovery cohort. In individuals who presented with clinical FTLD or FTLD-TDP, rs1990622 genotype was significantly associated with age at death in both univariate models and models adjusting for age and presence/absence of motor neuron disease (MND). In individuals who presented with ALS, rs1990622 genotype was not significantly associated with age at death, with a trend towards association with age at onset. Asterisks denote significance.

for three key reasons. First, since the FTLD-TDP GWAS predated the discovery of *C9orf72* expansions as a cause of FTLD, this large, international cohort was unbiased in enrollment with respect to *C9orf72* status. Second, all cases were neuropathologically confirmed to have FTLD-TDP, ensuring neuropathological homogeneity. Third, because all cases had undergone genome-wide genotyping, we could be certain that cases did not have cryptic familial relationships with one another.

As shown in **Table 3.2**, rs1990622 genotype was again correlated with age at death in this cohort (n=75), in both univariate analyses (*P*=0.016) and linear regression models adjusting for sex and the presence or absence of MND (*P*=0.019). Moreover, in this larger replication cohort, rs1990622 genotype was also correlated with age at onset (n=68 with age at onset data, *P*=0.019 for univariate analyses and *P*=0.032 for multivariate analyses adjusting for sex and presence or absence of MND). Consistent with the results from our discovery cohort, the minor allele (C) of

Disease	Outcome	Predictors	Beta (rs1990622, each minor allele)	R^2 for model	*P*-value (rs1990622)
FTLD-TDP	Age at Death (n=75)	rs1990622	-3.342	0.048	**0.016 ***
		rs1990622, Sex, MND	-3.413	0.032	**0.019 ***
	Age at Onset (n=68)	rs1990622		-3.473	
		rs1990622, Sex, MND		-3.198	

Table 3.2. *TMEM106B* genotype affects age at death and age at onset in *C9orf72* expansion carriers in a multi-site FTLD-TDP replication cohort.
Linear regressions were used to evaluate the effect of *TMEM106B* genotype at rs1990622 on the age at death or age at onset in *C9orf72*+ FTLD from a multi-site replication cohort of FTLD-TDP cases. Rs1990622 genotype was significantly associated with both age at death and age at onset, in both univariate models and models adjusting for age and presence/absence of motor neuron disease (MND). Asterisks denote significance.

rs1990622 was associated with earlier age at death, as well as earlier age at onset. Indeed, patients showed earlier disease onset and earlier death by more than three years for each additional minor allele at rs1990622 carried.

We further examined this genetic modifier effect using Kaplan-Meier survival analyses performed on the combined cohort (discovery plus replication, n=89 for age at death analysis, n=94 for age at onset analysis) of *C9orf72*+ FTLD cases. As shown in **Figure 3.1**, *TMEM106B* genotypes at rs1990622 were significantly associated with age at death (**Figure 3.1A**, *P*=0.046, log rank test for trend), with a trend towards association for age at onset (**Figure 3.1C**, *P*=0.064) in this combined cohort. In addition, we observed that the curve separation between rs1990622 minor allele homozygotes (CC) and heterozygotes (TC) was greater than the separation between heterozygotes (TC) and major allele homozygotes (TT). We therefore re-analyzed our data under a major-allele dominant model for rs1990622 and observed a stronger effect of *TMEM106B* genotype on age at death (p=0.041, log rank test for trend) and age at onset (*P*=0.037, log rank test for trend) in *C9orf72*+ FTLD. Indeed, at any given age, CC homozygotes at rs1990622 had more than twice the risk of manifesting disease (**Figure 3.1D**, HR 2.022, 95% CI 1.042-3.925), and

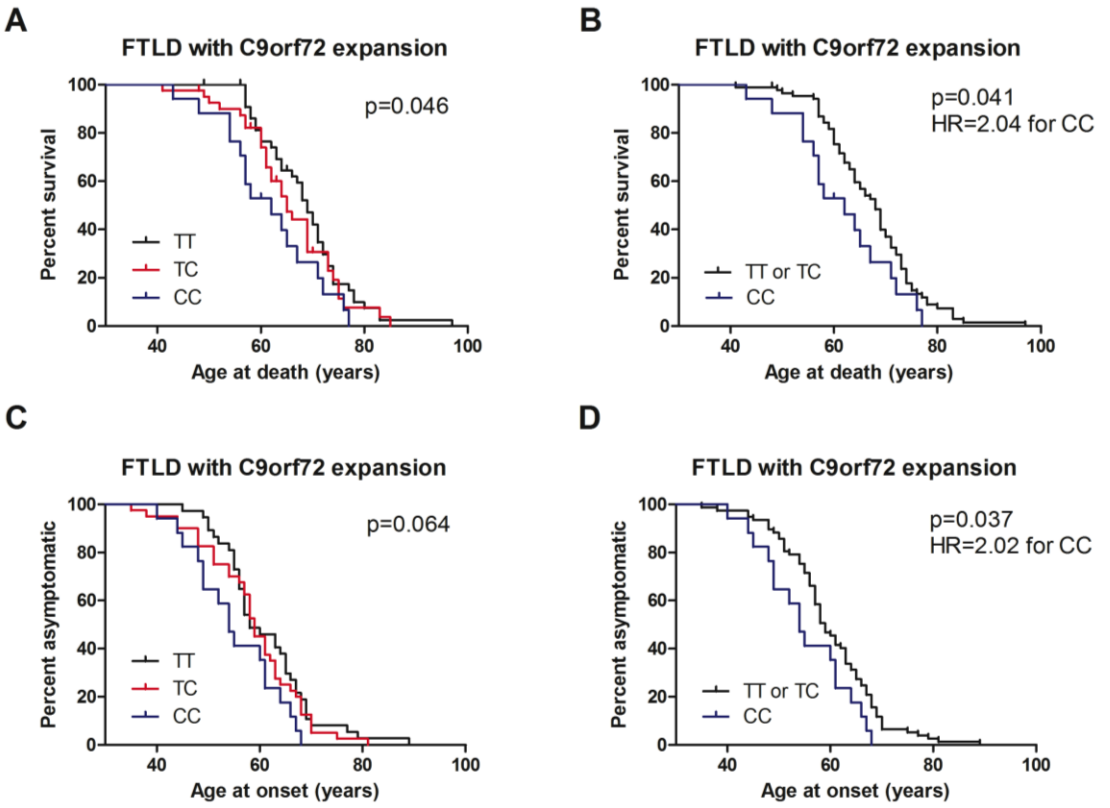


Figure 3.1. *TMEM106B* genotype influences age at death and age at onset in *C9orf72*+ FTLD.

All survival analyses were performed in 104 total *C9orf72*+ FTLD cases, from the combined discovery and replication cohorts. Of these 104 total cases, 89 had available age-at-death data, and 94 had age-at-onset data.

(A) Age at death was significantly associated with *TMEM106B* genotype at rs1990622, the top SNP associated with FTLD-TDP in our prior GWAS. Log rank test for trend two-tailed p=0.046, assuming a codominant model.

(B) Under a major-allele-dominant model, *TMEM106B* rs1990622 genotype was even more significantly associated with age at death, with more than twice the risk of death at any given age for CC carriers compared to carriers of one or more T alleles (two-tailed *P*=0.041, HR=2.039, 95% CI 1.031-4.033).

(C) Age at onset showed a trend towards association with *TMEM106B* genotype at rs1990622. Log rank test for trend two-tailed *P*=0.064, assuming a codominant model.

(D) Under a major-allele-dominant model, *TMEM106B* rs1990622 genotype showed a significant association with age at disease onset, with more than twice the risk of disease onset at any given age for CC carriers compared to carriers of one or more T alleles (two-tailed *P*=0.037, HR=2.022, 95% CI 1.042-3.925).

more than twice the risk of death (**Figure 3.1B**, HR 2.039, 95% CI 1.031-4.033), compared to other genotypes.

3.3.3 *TMEM106B* genotype does not exert a genetic modifier effect in *C9orf72* expansion negative FTLD-TDP cases

We next asked whether the *TMEM106B* genetic modifier effect observed for *C9orf72*+ FTLD extended to FTLD-TDP cases without *C9orf72* expansions, again using FTLD-TDP cases from the FTLD-TDP GWAS for which *C9orf72* and/or *GRN* mutation status was known. We considered cases with and without *GRN* mutations separately.

As shown in **Figure 3.2A**, *TMEM106B* rs1990622 genotype did not affect age at death in FTLD-TDP cases without *C9orf72* expansions or *GRN* mutations (n=241). In the subset of *GRN*-related FTLD-TDP (n=116, **Figure 3.2B**), only one rs1990622 CC individual had age at death information available, so we could only compare TT and TC individuals, who did not differ significantly in age at death. Similar results were obtained for age-at-onset analyses (data not shown).

3.3.4 *TMEM106B* genotype is associated with FTLD-TDP in *C9orf72* expansion carriers

The observed genetic modifier effect for *TMEM106B* in *C9orf72*+ FTLD is surprising in its direction. Specifically, the rs1990622 minor allele associated with decreased risk of FTLD-TDP by GWAS is correlated with younger age at onset and death among *C9orf72*+ FTLD cases, implying a deleterious effect in this mutation subgroup. We therefore examined *TMEM106B* rs1990622 allele frequencies in 116 G*RN*+ FTLD cases, 80 *C9orf72*+ FTLD cases, and 241 FTLD-TDP cases in which mutations in *GRN* and expansions in *C9orf72* had been excluded. As with the age-at-onset and age-at-death analyses, FTLD-TDP cases were from our prior FTLD-TDP GWAS, although numbers in each group are slightly higher because individuals with genotypes but lacking age-at-death or age-at-onset data could be included. As shown in **Table 3.3**, *TMEM106B* rs1990622 genotype was significantly associated with FTLD-TDP in all three subgroups, with the same direction of association. In each case, the major allele of rs1990622 was enriched in disease.

3.3.5 *TMEM106B* genotype is not associated with plasma progranulin levels in *C9orf72* expansion carriers

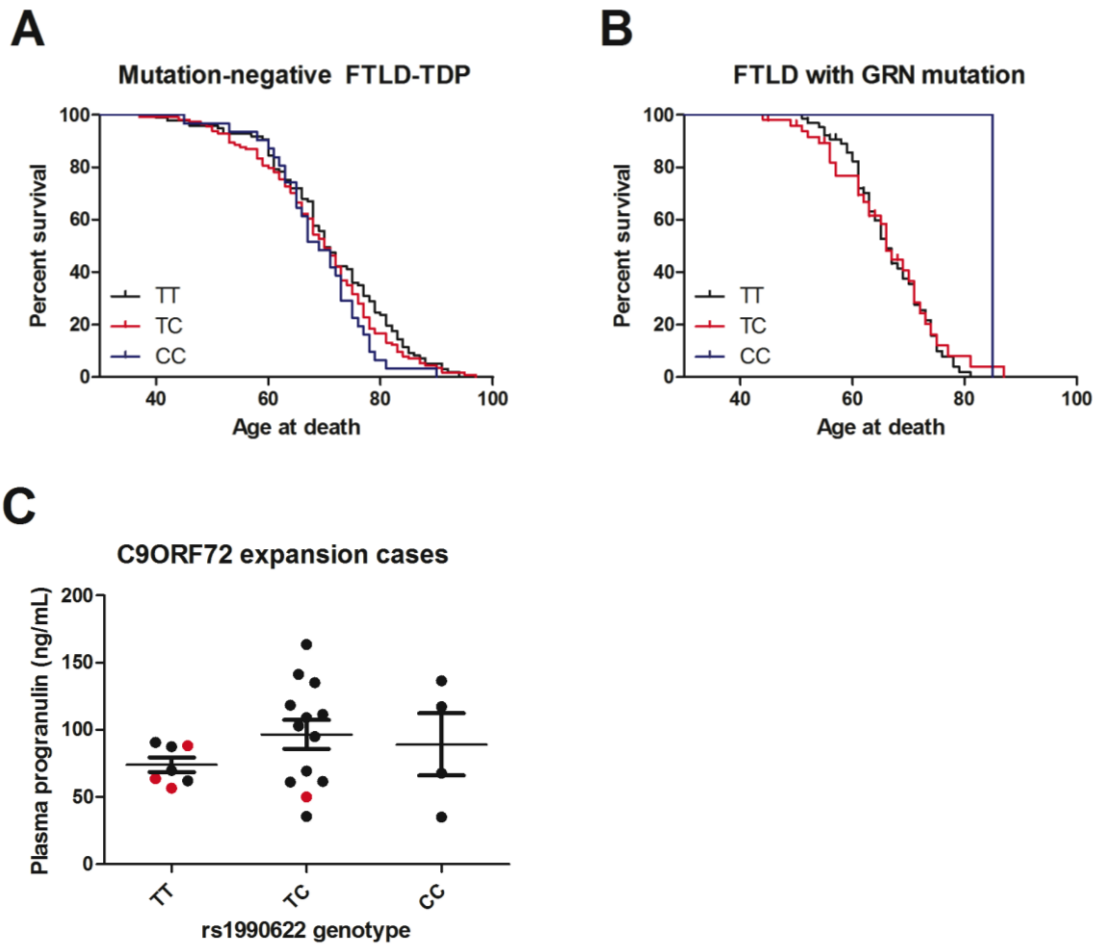


Figure 3.2 *TMEM106B* genotype does not affect age at death or age at onset for FTLD-TDP without *C9orf72* expansions.
(A) In 241 FTLD-TDP cases negative for *GRN* mutations or *C9orf72* expansions, *TMEM106B* genotype at rs1990622 did not affect age at death.
(B) In 116 FTLD-TDP cases with *GRN* mutations, we found no significant difference in age at death comparing TT and TC carriers at rs1990622. In this cohort, only one individual had the CC genotype, precluding our ability to evaluate the influence of this genotype.
(C) Plasma progranulin levels were measured in a convenience subset of 24 *C9orf72* expansion carriers by ELISA. Progranulin levels did not differ significantly by *TMEM106B* rs1990622 genotype, although the TT carriers exhibited significantly less variance in their progranulin levels. Black dots indicate individuals who presented with ALS, while red dots indicate individuals who presented with FTLD.

TMEM106B genotype has been reported to influence plasma progranulin levels in healthy individuals and *GRN*+ FTLD, with the rs1990622 minor allele associated with increased progranulin expression. We evaluated whether this relationship was also true in *C9orf72* expansion carriers.

Disease status	N	rs1990622 Major allele T	rs1990622 Minor allele C	*P*-value
Normal	2509	0.564	0.436	-
***GRN*+ FTLD-TDP**	116	0.776	0.224	**<0.0001**
***C9orf72*+ FTLD-TDP**	80	0.669	0.331	**0.008**
FTLD-TDP (no mutation)	241	0.640	0.360	**0.001**

Table 3.3. *TMEM106B* rs1990622 genotype is associated with FTLD-TDP in all genetic subgroups.
Chi-square tests were performed to evaluate for association between disease and rs1990622 genotype for FTLD-TDP subgroups defined by the presence of *GRN* mutations (*GRN*+ FTLD-TDP), presence of *C9orf72* expansions (*C9orf72*+ FTLD-TDP), or the absence of both genetic mutations (FTLD-TDP (no mutation)). The major allele was significantly associated with disease in all three subgroups. Allele frequencies for normal controls provided here are from our previously published GWAS.

In a convenience subset of 24 *C9orf72* expansion carriers (20 with *C9orf72*+ ALS and 4 with *C9orf72*+ FTLD) from the UPenn discovery cohort for whom we had plasma samples, we measured progranulin levels using an enzyme-linked immunosorbent assay (ELISA). As shown in **Figure 3.2C**, there were no significant differences in plasma progranulin levels comparing *C9orf72* expansion carriers with TT, TC, and CC genotypes at rs1990622. Adjusting for sex and age at plasma sampling or duration of disease did not affect this result. Additionally adjusting for clinical manifestation as FTLD or ALS did not affect this result.

3.4 DISCUSSION

In the current study, we find that *TMEM106B* is a genetic modifier for *C9orf72*+ FTLD, demonstrating a significantly earlier age at death for *TMEM106B* rs1990622 minor allele (C) carriers. This effect appears to be specific to *C9orf72*+ FTLD, since *C9orf72*-expansion-negative FTLD cases do not differ in age at death depending on rs1990622 genotype. Finally, among *C9orf72* expansion carriers, we do not see a clear effect of rs1990622 genotype on plasma progranulin levels.

We observe that *TMEM106B* genotypes exert a genetic modifier effect in *C9orf72*+ FTLD. Examples of common risk variants acting as genetic modifiers in Mendelian subgroups of disease are increasingly being described. In the field of neurodegeneration, one well-known example is the age-at-onset modifying effect of Apolipoprotein E (*APOE*) isoform in *PSEN2*-related-Alzheimer's Disease (Wijsman et al., 2005). Moreover, in *GRN*+ FTLD, *TMEM106B* has been reported as a genetic modifier affecting both age-at-onset and circulating levels of progranulin (Cruchaga et al., 2011; Finch et al., 2011).

What is more unusual in this case is the direction of the genetic modifier effect. Specifically, the *TMEM106B* allele that is associated with decreased risk of developing FTLD-TDP (Van Deerlin et al., 2010) (and later age at onset in *GRN*+ FTLD (Cruchaga et al., 2011)) appears to accelerate the disease phenotype (associating with earlier age at death and onset) in *C9orf72*+ FTLD. This effect may be an example of the general phenomenon of sign epistasis, in which a genetic variant is beneficial on some genetic backgrounds but deleterious in others. Sign epistasis has been demonstrated in lower organisms such as bacteria (Silva et al., 2011), but may certainly generalize to higher organisms as well (Kern and Kondrashov, 2004; Kondrashov et al., 2002). In the few reported empirically-derived examples of sign epistasis, the two (or more) genetic loci involved converge mechanistically in, for example, antibiotic resistance pathways (Schenk et al., 2013) or enzyme-substrate interactions (Zhang et al., 2012). Thus, the observed epistasis between *TMEM106B* and *C9orf72* suggests that these two proteins may have convergent functions in the pathophysiology of FTLD-TDP. Intriguingly, TMEM106B has been linked to endosomal-lysosomal

pathways (Brady et al., 2013; Chen-Plotkin et al., 2012; Lang et al., 2012; Nicholson et al., 2013). The largely uncharacterized protein C9orf72 is structurally related to DENN protein family members (Levine et al., 2013). DENN proteins function in the regulation of Rab GTPases, which in turn regulate the many membrane trafficking events needed for proper function of the endosomal-lysosomal pathway.

We note that *TMEM106B* rs1990622 genotypes differ in allelic frequencies between *C9orf72*+ FTLD-TDP and normal controls; this situation in which a common variant shows allelic association with disease even in a monogenic, highly-penetrant subgroup of disease has been reported in *GRN*+ FTLD-TDP as well (Finch et al., 2011; Van Deerlin et al., 2010). In the case of the *GRN* mutants, a potential explanation may lie in ascertainment bias, since *TMEM106B* risk variant carriers may manifest disease at an earlier age (Cruchaga et al., 2011), making it more likely for them to be included in a cross-sectional sampling of diseased individuals. Such an argument cannot explain our current result, however, since the rs1990622 major allele (found by genome-wide association to be enriched in FTLD-TDP) appears to delay age at death and age at onset in *C9orf72*+ FTLD cases. An alternate explanation may lie in the fact that *C9orf72* expansions have a broad range of phenotypic expression, manifesting as ALS, FTLD, or a syndrome combining both motor neuron disease and dementia. We have previously shown that ALS patients who are major allele carriers at rs1990622 are more likely to demonstrate cognitive impairment (Vass et al., 2011). Thus, it is possible that *TMEM106B* genotype modulates the phenotypic expression of *C9orf72* expansions, with rs1990622 major allele carriers more likely to manifest clinically with dementia. Whether an effect of directing regional pathology towards cognitive regions rather than motor regions also underlies the apparently protective effect on age at death for *TMEM106B* rs1990622 major allele carriers with *C9orf72* expansions remains to be seen.

It is notable that we were able to replicate the genetic modifier effect of *TMEM106B* genotype in *C9orf72*+ FTLD in a 30-site, international cohort of subjects. Undoubtedly, site-to-site variation in methods of ascertaining age at onset would contribute to noise, and site-to-site variation

in practice with respect to aggressiveness of clinical care with a fatal neurodegenerative disease would contribute to differences in age at death in such a dataset. The ability to see a significant epistatic effect of *TMEM106B* on *C9orf72* in such a cohort, nonetheless, may have been helped by the fact that our replication cohort was homogeneous with respect to neuropathology (all FTLD-TDP) and genome-wide genotyping in these individuals allowed us to exclude individuals with cryptic familial relationships that might cloud the picture. In any case, the international, multi-site nature of our replication cohort increases our confidence that our findings are not due to artifact.

The current study has several limitations. First, while we did not see an age-at-death-modifying effect for *TMEM106B* in *C9orf72* expansion-associated ALS, our sample size was small (n=39) and likely underpowered to adequately address this question. Thus, future studies examining this relationship in more *C9orf72*-expansion-related ALS cases would be a valuable addition to the data presented here. Second, we did not see a clear modifier effect of *TMEM106B* genotype in the *GRN*+ FTLD-TDP cases in this study, as has been previously reported (Cruchaga et al., 2011). However, our study had only one rs1990622 minor allele homozygote in the *GRN*+ FTLD subgroup, precluding our ability to examine *TMEM106B* genotype effect in a major-allele-dominant model. Third, we were able to obtain plasma samples on 24 *C9orf72* expansion carriers, in whom we measured progranulin levels. Plasma progranulin levels did not differ by *TMEM106B* genotype in this set of samples, which could reflect either insufficient sample size or a biologically-relevant finding. Should further studies in larger sample sizes corroborate our result, this would suggest that *C9orf72* expansions may interrupt the means by which TMEM106B affects circulating progranulin levels.

In conclusion, we demonstrate here that *TMEM106B* is the first reported genetic modifier in *C9orf72* expansion-related FTLD. Our findings suggest a previously unsuspected link between these two proteins in the pathophysiology of FTLD and open up new directions for the development of disease-modifying therapy.

3.5 METHODS

Patient cohorts

FTLD and ALS cases with *C9orf72* expansions of greater than 30 hexanucleotide repeats were identified from among cases in the Integrated Neurodegenerative Disease Database at the University of Pennsylvania (UPenn) to form a discovery cohort (Toledo et al., 2013; Xie et al., 2011). Patients were initially seen at the UPenn Frontotemporal Degeneration Center (FTDC), Amyotrophic lateral sclerosis Center (ALSC), or Alzheimer's Disease Center (ADC); all were collected with Institutional Review Board Approval. In addition to having a *C9orf72* expansion, the criteria for selection of FTLD cases was a pathological diagnosis of FTLD-TDP (n=10) or a clinical diagnosis of FTLD or FTLD-MND (n=19), according to published criteria (Gorno-Tempini et al., 2011; Litvan et al., 1996; Mackenzie et al., 2011; McKhann et al., 2001; Rascovsky et al., 2011; Strong et al., 2009). *C9orf72*+ ALS cases (n=55) all met El Escorial-revised criteria (Brooks et al., 2000). Twenty of the 55 ALS cases had autopsy confirmation of ALS pathology. For both FTLD and ALS cases, only probands were selected. In situations where patients exhibited both dementia and motor neuron disease (MND), cases were assigned to FTLD-MND if the initial presentation was cognitive and to ALS if the initial presentation was MND. All *C9orf72*+ FTLD and *C9orf72*+ ALS cases meeting these criteria were included without bias for familial-vs.-apparently-sporadic patterns of inheritance, and without prior knowledge of *TMEM106B* genotype.

The *C9orf72*+ FTLD cohort is 93.5% white (6.5% unknown ethnicity) and 54.8% male. The *C9orf72*+ ALS cohort is 87.2% white, 5.6% black, 3.5% Latino, and 3.7% unknown ethnicity with 59.8% males. Age at onset and age at death were collected, but both were not available on all subjects (*e.g.* no age at death for living subjects, and sometimes no known age at onset for autopsy cases), therefore the numbers of cases from each cohort vary depending on the data needed for analysis.

The previously published and publicly available FTLD-TDP GWAS from the International Collaboration for Frontotemporal Lobar Degeneration was used as a replication cohort (Van Deerlin et al., 2010). A subset of the FTLD-TDP cases were known from the original study to have a

pathogenic *GRN* mutation (n=116) and is used here as a comparison group (Chen-Plotkin et al., 2012; Van Deerlin et al., 2010). Cases lacking a *GRN* or *VCP* mutation (n=421) were screened for *C9orf72* expansions either by the contributing site (n=199) or by UPenn (n=142), using published methods (DeJesus-Hernandez et al., 2011; Renton et al., 2011). Some cases (n=84) were presumed to lack an expansion due to a BB genotype at rs2814707 and were not tested to conserve resources (Strong et al., 2009). 80 FTLD-TDP cases with *C9orf72* expansions were identified from 30 clinical sites that agreed to collaborate on this project. Contributing sites that provided *C9orf72* genetic data included: Erasmus University, Rotterdam, The Netherlands; Indiana University, Indianapolis, Indiana; Banc de Teixits Neurologics-Biobanc-Hospital Clinic-IDIBAPS, Barcelona, Spain; Kings College, London, UK; UCL Institute of Neurology, Queen Square, London, UK; Ludwig-Maximilians University, Munich, Germany; University of New South Wales, Sydney, Australia; VIB, University of Antwerp, Antwerp, Belgium; Massachusetts General Hospital, Boston, Massachusetts; University of Sheffield, Sheffield, UK; Institut National de la Santé et de la Recherche Laboratoire de Neuropathologie, Paris, France. Contributing sites with *C9orf72*+ cases identified at UPenn included: Sydney Brain Bank, Australia; Boston University, Boston, Massachusetts; Duke University, Durham, North Carolina; Emory University, Atlanta; Georgia; Karolinska Institute, Stockholm, Sweden; Mt. Sinai School of Medicine, Bronx, New York; Oregon Health Sciences University, Portland, Oregon; University of Pittsburgh, Pittsburgh, Pennsylvania; Rush University, Chicago, Illinois; University of Texas Southwestern, Dallas, Texas; University of Toronto, Toronto, Canada; University of California (Davis, Irvine, San Diego campuses), California; University of Michigan, Ann Arbor, Michigan; University of Kuopio, Finland; University of Southern California, Los Angeles, California; Washington University, St. Louis, Missouri; University of Pennsylvania, Philadelphia, Pennsylvania. Of the 80 cases, 5 UPenn cases overlapped with the UPenn discovery cohort and were removed, leaving 75 *C9orf72* expansion cases for analysis in the replication cohort. 325 cases remained after identification of *GRN* and *VCP* mutation carriers, and *C9orf72* expansion carriers. Of these, 241 cases had been formally tested for (and found negative for) *C9orf72* expansions, and these were used as the mutation-negative FTLD-TDP

cohort; an additional 84 cases may also be negative for *C9orf72* expansions (due to their genotype at rs2814707), but since they were not formally tested, they were excluded from analysis.

Genotyping

DNA from UPenn cases, extracted from blood or brain samples as previously described (Van Deerlin et al., 2010),was tested for rs1990622 genotype using one of two methods: TaqMan chemistry-based allelic discrimination assays as previously described (Chen-Plotkin et al., 2012; Van Deerlin et al., 2010), or a custom Sequenom MassArray genotyping panel that includes PCR and extension primers for rs1990622. PCR and extension primer sequences for the Sequenom panel are available on request. Both genotyping methods were compared and found to be concordant (data not shown) (Toledo et al., 2013).

Plasma progranulin measurement

Plasma samples were collected from UPenn ALS and FTLD discovery cohort patients, aliquotted, and stored at -80°C as previously described (Chen-Plotkin et al., 2011). Progranulin levels were measured using a commercially available sandwich ELISA (Human progranulin ELISA kit, AdipoGen), according to manufacturer instructions.

Statistical analyses

Linear regression analyses evaluating the association of *TMEM106B* genotype with age at death or age at disease onset were performed in R, with or without covariates as described in the text. Two-tailed p-values are reported for the discovery cohort, and one-tailed p-values are reported for the FTLD-TDP GWAS replication cohort, since the expected directionality was known. For the combined dataset, survival analyses (Kaplan-Meier method) were also performed in Prism, and two-tailed p-values from the log-rank test for trend are reported.

Where indicated, codominant, major-allele-dominant, and minor-allele dominant models of genetic effect were investigated.

In addition, we tested for association between *TMEM106B* genotype and disease for genetically-defined subsets of FTLD (*C9orf72*+ FTLD, *GRN*+ FTLD, or individuals without *C9orf72*

expansions or *GRN* mutations). Chi-square statistics were calculated for rs1990622 using the FTLD-TDP GWAS cases and controls (Van Deerlin et al., 2010).

For plasma progranulin analyses, Kruskal-Wallis tests were used to compare plasma progranulin measures among carriers of different *TMEM106B* genotypes under a codominant model, and Mann-Whitney tests were used to compare different *TMEM106B* genotypes under major-allele-dominant and minor-allele dominant models. In addition, multivariate linear regressions predicting plasma progranulin levels from *TMEM106B* genotype were used to adjust for sex, age, duration of disease, or clinical manifestation as described in the text.

R-scripts for analyses are available upon request.

INTERNATIONAL COLLABORATION FOR FRONTOTEMPORAL LOBAR DEGENERATION

The International Collaboration for Frontotemporal Lobar Degeneration consisted of 45 clinical sites collaborating to collect cases for an FTLD-TDP genome-wide association study (GWAS); this GWAS led to the discovery that common variants in *TMEM106B* are a genetic risk factor for FTLD-TDP. Members of the Collaboration who contributed *C9orf72*+ FTLD-TDP cases for this study include Irina Alafuzoff, Anna Antonell, Nenad Bogdanovic, William Brooks, Nigel Cairns, Johnathan Cooper-Knock, Carl W. Cotman, Patrick Cras, Marc Cruts, Peter P. De Deyn, Charles DeCarli, Carol Dobson-Stone, Sebastiaan Engelborghs, Nick Fox, Douglas Galasko, Marla Gearing, Ilse Gijselinck, Jordan Grafman, Paivi Hartikainen, Kimmo J. Hatanpaa, J. Robin Highley, John Hodges, Christine Hulette, Paul G. Ince, Lee-Way Jin, Janine Kirby, Julia Kofler, Jillian Kril, John J. B. Kwok, Allan Levey, Andrew Lieberman, Albert Llado, Jean-Jacques Martin, Eliezer Masliah, Christopher J. McDermott, Catriona McLean, Ann C. McKee, Simon Mead, Carol A. Miller, Josh Miller, David Munoz, Jill Murrell, Henry Paulson, Olivier Piguet, Martin Rossor, Raquel Sanchez-Valle, Mary Sano, Julie Schneider, Lisa Silbert, Salvatore Spina, Julie van der Zee, Tim Van Langenhove, Jason Warren, Stephen B. Wharton, Charles L. White III, Randall Woltjer.

CHAPTER 4: A DEMENTIA-ASSOCIATED RISK VARIANT NEAR TMEM106B AFFECTS CHROMATIN ARCHITECTURE AND GENE EXPRESSION

by

Michael D. Gallagher, Marijan Posavi, Peng Huang, Travis L. Unger, Yosef Berlyand, Analise L. Gruenewald, Alessandra Chesi, Elisabetta Manduchi, Andrew D. Wells, Struan F. Grant, Gerd A. Blobel, Christopher D. Brown, and Alice S. Chen-Plotkin

This chapter was written by Michael Gallagher and Alice Chen-Plotkin, with input from the other authors. The majority of the experiments and statistical analyses performed in this chapter were performed by Michael Gallagher, whereas the downstream high-throughput sequencing analyses were performed by Marijan Posavi.

4.1 ABSTRACT

Neurodegenerative diseases pose an extraordinary threat to the world's aging population; unfortunately, no disease-modifying therapies are available. While genome-wide association studies (GWAS) have identified hundreds of novel risk loci for neurodegeneration, the mechanisms by which these loci influence disease risk are largely unknown. Here, we show that common genetic variants associated with risk for frontotemporal dementia correlate with increased expression of the gene *TMEM106B*. Furthermore, incremental increases in *TMEM106B* expression result in incremental increases in cell toxicity. Using both bioinformatic and bench-based approaches, we find that a noncoding variant, rs1990620, influences CTCF-mediated long-range interactions between distal regulatory elements, suggesting a causal mechanism for allele-specific expression and disease association at this locus. We further show that genetic variants associated with risk for neurodegenerative diseases beyond frontotemporal dementia are enriched in brain CTCF-binding sites genome-wide, implicating CTCF-mediated gene regulation in risk for neurodegeneration more generally.

4.2 INTRODUCTION

Neurodegenerative diseases are a leading cause of disability and death in the developed world, with numbers affected by these diseases poised to increase as the world population ages. There are still no disease-modifying therapies for the major late-onset neurodegenerative diseases such as Alzheimer's disease (AD), Parkinson's disease (PD), frontotemporal lobar degeneration, and amyotrophic lateral sclerosis (Chen and Zheng, 2012). To generate novel leads in tackling this growing problem, many genome-wide association studies (GWAS) have been performed in the various neurodegenerative diseases, involving >100,000 patients, and identifying >200 genetic risk loci (Welter et al., 2014). While genetic risk loci have been utilized, singly or in aggregate, to refine predictions for risk of developing disease (Abraham and Inouye, 2015; Nalls et al., 2014), the greatest potential for these GWAS-identified loci may lie in the identification of novel disease mechanisms (Ramanan and Saykin, 2013).

However, the interpretation of disease-associated risk loci is complicated. The "sentinel" variant, usually a single nucleotide polymorphism (SNP) identified by GWAS, is rarely the specific change in DNA sequence – or "causal" variant – that results at the molecular level in a mechanistic change. Instead, in most cases, tens or hundreds of genetic variants at each locus are in strong linkage disequilibrium (LD) with the sentinel variant, constituting a set of co-inherited variants – or haplotype – any of which may be the underlying cause for increased disease risk (Edwards et al., 2013). Indeed, the risk-associated haplotype may span multiple genes, making even the gene to which a GWAS signal belongs unclear. Given these complexities, it is perhaps unsurprising that, with one exception pertaining to common variants near the *SNCA* gene, which was already implicated prior to the GWAS era in the development of PD (Soldner et al., 2016), none of the neurodegenerative disease risk loci identified by GWAS have been characterized in molecular detail. Yet such a molecularly precise understanding of a GWAS-identified genetic risk locus is a likely prerequisite for downstream therapeutic development.

Frontotemporal lobar degeneration (FTLD) is a neurodegenerative dementia affecting ~10-20 per 100,000 persons between the ages of 45 and 64, making FTLD the second most common

early-onset dementia (Bang et al., 2015; Seelaar et al., 2011). FTLD is a fatal, untreatable disease, with death typically occurring within 8 years after diagnosis (Bang et al., 2015). Noncoding single nucleotide polymorphisms (SNPs) on chromosome 7p21 have been associated with risk for the major neuropathological subtype of FTLD, characterized by pathological inclusions of the protein TDP-43 (FTLD-TDP) (Van Deerlin et al., 2010). The association of this locus with FTLD-TDP has been replicated (Finch et al., 2011; Hernandez et al., 2015; van der Zee et al., 2011), and the major T allele of the sentinel SNP, rs1990622, yielded an odds ratio of ~1.6 for disease development (Van Deerlin et al., 2010). Genotype at rs1990622 also affects age at disease onset in Mendelian forms of FTLD-TDP (Cruchaga et al., 2011; Finch et al., 2011; Gallagher et al., 2014), as well as risk for development of cognitive impairment in the related disorder amyotrophic lateral sclerosis (ALS) (Vass et al., 2011). We and others have implicated a gene in this region, *TMEM106B*, as being causal (Brady et al., 2013; Chen-Plotkin et al., 2012; Lang et al., 2012). However, studies to date have not explained how genetic variation at the 7p21 locus affects the function of *TMEM106B* or another gene, thereby contributing to the pathogenesis of FTLD-TDP.

In this study, we demonstrate that (1) common GWAS-implicated variants associated with FTLD-TDP are correlated with expression levels of *TMEM106B*, with increased expression correlating with the risk haplotype, (2) incremental increases in *TMEM106B* expression are associated with incremental increases in cell toxicity, (3) the risk allele of a candidate causal variant (rs1990620) in complete LD with rs1990622, the GWAS sentinel SNP, increases recruitment of the chromatin organizing protein CCCTC-binding factor (CTCF), and (4) long-range chromatin looping interactions at the *TMEM106B* locus are stronger on the chromosome bearing the risk-associated allele. Together, these data provide a molecularly detailed mechanism for the effect of common genetic variation at this locus on risk for neurodegenerative disease.

4.3 RESULTS

4.3.1 Genetic variation at the 7p21 locus associates with TMEM106B expression

It is increasingly recognized that many GWAS-implicated variants associated with disease risk may confer their effects by altering the expression levels of nearby genes (Maurano et al., 2012; Schaub et al., 2012). In the case of 7p21, several studies have demonstrated such an expression quantitative trait locus (eQTL) effect for *TMEM106B* in multiple human tissue types, including brain and Epstein-Barr virus (EBV)-immortalized B lymphoblastoid cell lines (LCLs) (Dixon et al., 2007; Liang et al., 2013; Stranger et al., 2012; Yu et al., 2015). We therefore systematically investigated the 7p21 locus for all eQTL effects, in order to confirm the *TMEM106B* eQTL effect and to exclude other potential causal genes at this locus.

Analysis of data from The Genotype-Tissue Expression (GTEx) Project (GTEx Consortium, 2015) demonstrated a robust association between genotype at rs1990622, the GWAS sentinel SNP, and *TMEM106B* expression in several cell types from normal individuals (**Figure 4.1A**), including LCLs (n=114, P=1.9x10^{-6}) and transformed fibroblasts (n=272, P=3.0x10^{-7}). No other transcript genome-wide was significantly associated with rs1990622 genotype. The association between rs1990622 and *TMEM106B* levels has also been previously reported in human brain (Yu et al., 2015); GTEx data confirms this relationship in the hippocampal and nucleus accumbens brain regions (**Figure 4.1B**). In all cell types, the risk allele was consistently associated with increased *TMEM106B* expression.

The 7p21/*TMEM106B* locus is harbored within a 36kb LD block in samples from individuals of European ancestry, the population in which the original FTLD-TDP GWAS was performed (**Figure 4.1C**). This LD block encompasses the *TMEM106B* promoter, the entirety of the *TMEM106B* gene, and extends ~10kb downstream of the gene. According to 1000 Genomes data (1000 Genomes Project Consortium et al., 2015), the block contains 104 genetic variants that are in strong, but not perfect, LD with rs1990622 ($r^2 \geq 0.8$, **Figure 4.1C**). Indeed, in-depth examination of the eQTL effect in human hippocampal (**Figure 4.1D**) or LCL (**Figure S4.1**) samples from GTEx reveals multiple variants in strong LD with rs1990622 that are associated with *TMEM106B*

expression to a similar degree. We thus asked whether more than one eQTL signal occurs in this region, whether the eQTL association can be distinguished from association with FTLD-TDP, and what the candidate causal variant(s) underlying association with disease or expression might be.

We began by honing the region of eQTL association. To do this, we performed a second eQTL analysis of LCLs from eight ethnic populations (Stranger et al., 2012), reasoning that the different haplotype structures seen in disparate populations might refine the 36kb LD block of association seen in individuals of European ancestry. We found that with the addition of these populations, the region of association with *TMEM106B* expression could be truncated on the 5' end, effectively removing the promoter and first two exons of the gene, and reducing the number of potential causal variants to 84 (75 SNPs and 9 indels, **Figure 4.1E**).

We then performed conditional analyses using the refined region of association from the multi-ethnic analysis (Stranger et al., 2012). Conditioning on either the GWAS sentinel SNP, rs1990622, or the most significant eQTL SNP, rs6948844 (r^2=1 with rs1990622), yielded no variants within a 2Mb region that demonstrated any residual association with *TMEM106B* expression (**Figure 4.1F**), demonstrating that there is only one eQTL signal at this locus, and that the causal variant underlying association with *TMEM106B* expression may be the same as the causal variant underlying association with disease.

4.3.2 Increased levels of TMEM106B expression correlate with increased cellular toxicity

If the causal variant responsible for association with *TMEM106B* expression levels confers risk for neurodegeneration, one would expect incremental changes in *TMEM106B* expression to lead to incremental effects on cellular health. We and others have previously shown that over-expression of *TMEM106B* results in the development of enlarged LAMP1+ late endosomes/lysosomes appearing as vacuolar structures in multiple cell types, including neurons, with associated impairment in lysosomal degradative function (Brady et al., 2013; Busch et al., 2016; Chen-Plotkin et al., 2012; Stagi et al., 2014). However, the magnitudes of reported eQTL effects in human tissues are often modest (Dimas et al., 2009; GTEx Consortium, 2015), and so we sought to understand the effects of incremental increases in *TMEM106B* expression on

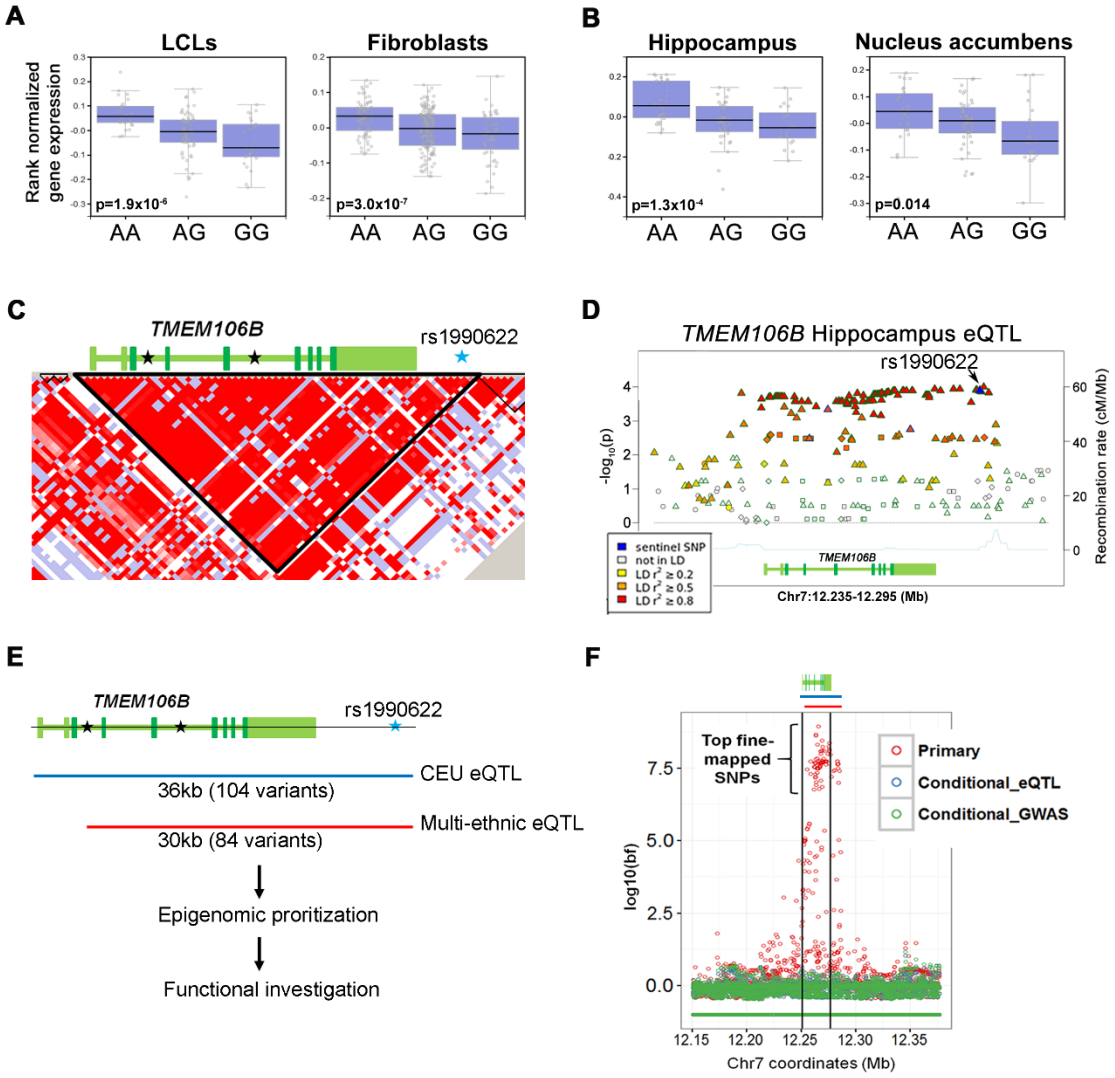


Figure 4.1. Analysis of expression quantitative trait locus (eQTL) effect at *TMEM106B*.
(A,B) Boxplots from the Genotype-Tissue Expression (GTEx) Project data demonstrate association of *TMEM106B* levels with rs1990622 genotype (A=risk allele) in peripheral cell types **(A)** as well as human brain regions **(B)**. Data from LCLs (n=114), fibroblasts (n=272), hippocampus (n=81) and nucleus accumbens (n=93) are shown. Black lines indicate median expression, lower and upper bounds of boxes indicate 25th and 75th percentile expression levels, respectively, and circles outside whiskers denote outliers.
(C) LD structure at *TMEM106B* in Caucasians, with gene structure indicated above LD plot. Coding exons are in dark green, UTRs are in light green, and SNPs associated by GWAS with frontotemporal lobar degeneration (FTLD) are indicated with stars, including the sentinel SNP, rs1990622.
(D,E) The *TMEM106B* eQTL effect extends across the 36kb LD block (shown in **(D)** for GTEx hippocampus data, with GRCh37/hg19 coordinates) that can be truncated on the 5' end with the addition of ethnicities with different haplotype structures **(E)**.
(F) Multi-ethnic conditional eQTL analysis of *TMEM106B* locus. Each circle represents a SNP, with genomic position on the x-axis and association with *TMEM106B* levels on the y-axis (log_{10}-transformed Bayes factor). *TMEM106B* gene and regions of eQTL association are indicated above

the plot. The primary multi-ethnic analysis (red) demonstrates a strong association between a SNP cluster and *TMEM106B* expression. Conditioning this analysis on either the top eQTL SNP (blue) or the sentinel GWAS SNP (green) results in loss of an association signal at this locus (*i.e.* there are no highly-associated SNPs shown in blue or green), suggesting that there is only one eQTL signal and that the causal variant underlying association with *TMEM106B* expression may be the same as the causal variant underlying association with FTLD.

disease-relevant measures such as (1) development of the previously-reported vacuolar phenotype and (2) cell toxicity.

To do this, we employed three different *TMEM106B* constructs that reliably produced a spectrum of over-expression ranging from ~2X to ~20X (**Figures 4.2A and 4.2B**). In HeLa cells, we found that with each incremental increase in *TMEM106B* expression over baseline, the percentage of cells exhibiting the vacuolar phenotype of enlarged lysosomes increased (**Figures 4.2C and 4.2D**), along with increased cell death (**Figure 4.2E**). Together, these findings suggest that genetic variation at the 7p21 locus may influence risk for neurodegeneration by altering *TMEM106B* expression-dependent effects on lysosomal function and cell health.

4.3.3 A candidate causal regulatory region

We next sought to identify the causal variant or variants responsible for allele-specific regulation of *TMEM106B* expression and, by extension, risk for FTLD-TDP. *TMEM106B* steady-state transcript levels depend on both the production of mRNA and its stability. We first considered the possibility of differential mRNA stability. In multiple LCLs homozygous at the *TMEM106B* locus, we found that mRNA stability did not differ between risk haplotype homozygotes and protective haplotype homozygotes (**Figure S4.2A**), suggesting that differences in the production of mRNA account for the observed eQTL effect.

To identify variants that may have transcriptional regulatory effects, we examined the 84 candidate variants (75 SNPs and 9 indels) located within the refined region of eQTL association (**Figures 4.1E and 4.1F**). We used ENCODE (ENCODE Project Consortium et al., 2012) and NIH Roadmap Epigenome (Roadmap Epigenomics Consortium et al., 2015) data to determine whether each variant is located in a predicted LCL *cis*-regulatory element (CRE), as determined by DNase

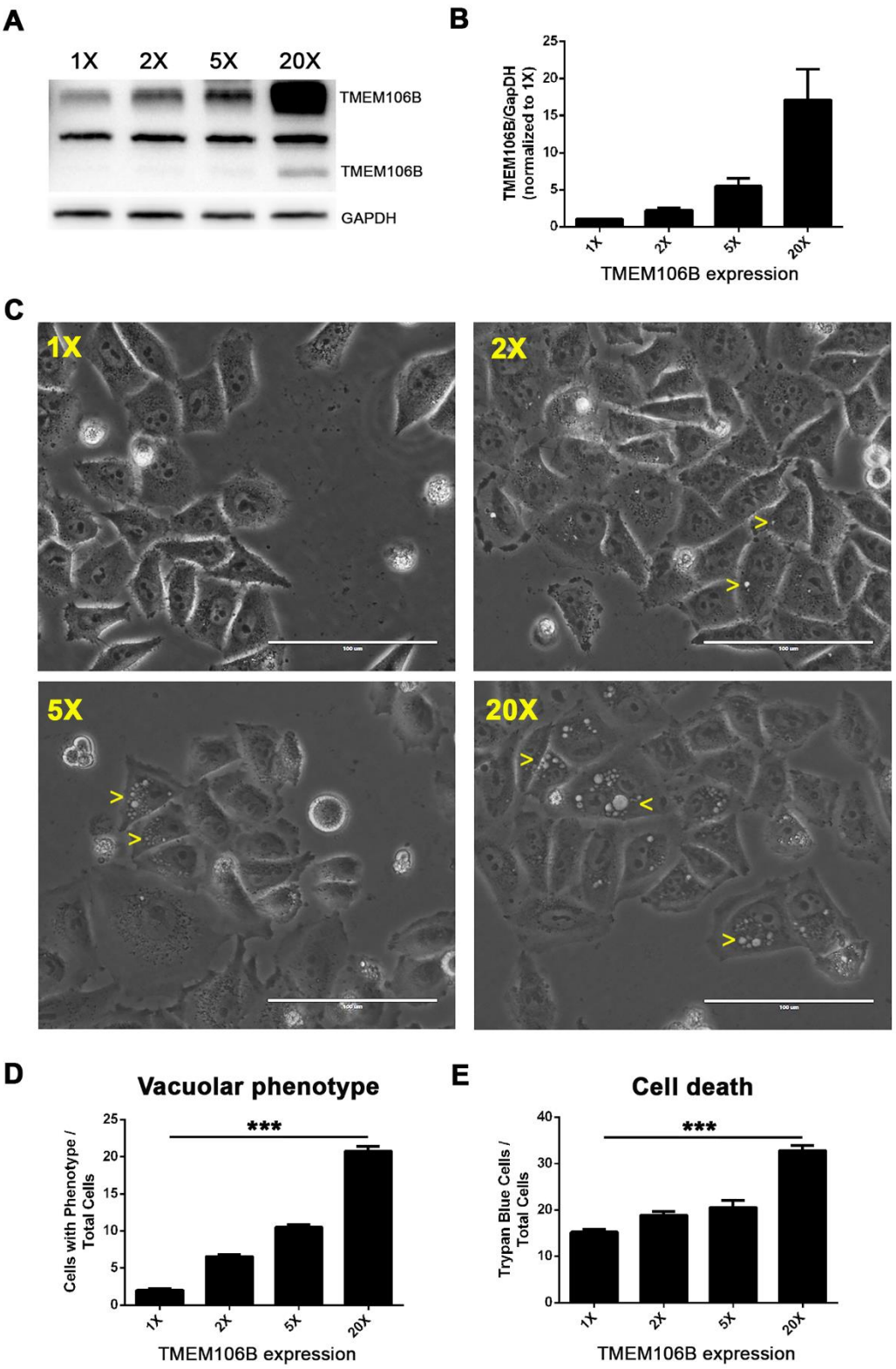

A
1X
2X
5X
20X
TMEM106B
TMEM106B
GAPDH
B
TMEM106B/GapDH
(normalized to 1X)
25
20
15
10
5
0
1X
2X
5X
20X
TMEM106B expression
C
1X
2X
5X
20X
D
Vacuolar phenotype
***
Cells with Phenotype /
Total Cells
25
20
15
10
5
0
1X
2X
5X
20X
TMEM106B expression
E
Cell death
***
Trypan Blue Cells /
Total Cells
40
30
20
10
0
1X
2X
5X
20X
TMEM106B expression

Figure 4.2. TMEM106B expression levels show dose-dependent effects on cell toxicity.
(A) Representative Western blot of TMEM106B levels in the absence (1X) and presence (2X, 5X, 20X) of various *TMEM106B* over-expression constructs transfected into HeLas. The bands at ~75kD and ~40kD represent dimeric and monomeric forms of TMEM106B. A non-specific band is indicated by the asterisk.
(B) Quantification of blots from six independent experiments (+/- SEM) demonstrate reliable expression levels of each construct.
(C) Representative bright-field images demonstrate a dose-dependent vacuolar phenotype in cells. Yellow arrowheads indicate cells exhibiting the phenotype.
(D,E) Quantification of the number of cells exhibiting **(D)** the vacuolar phenotype and **(E)** cell death is shown for each expression paradigm, across three independent experiments. Asterisks denote statistical significance ($p<0.001$ by ANOVA).

I hypersensitivity (DHS), transcription factor binding, or the active histone marks H3K27ac, H3K4me1 or H3K4me3.

Surprisingly, only seven SNPs spanning three candidate regulatory regions met these permissive conditions (**Figure 4.3 and Tables S4.1 and S4.2)**. Three SNPs located in intron 4 of *TMEM106B* overlap a region of DHS, transcription factor binding, and the enhancer-associated histone mark H3K4me1 in LCLs, while a fourth SNP downstream of *TMEM106B* overlaps a binding site for the transcription factor PU.1 in LCLs. Upon empirical testing in luciferase reporter assays, however, these regions displayed little or no enhancer activity in LCLs. Furthermore, the risk and protective haplotype versions of these regions did not differ in activity, suggesting that none of the overlapping SNPs affect regulatory activity (**Figures S4.2B-D**). The remaining candidate regulatory region contains the remaining three completely linked SNPs, one of which is rs1990622, the GWAS sentinel SNP. Moreover, in virtually all ENCODE-tested cell types, including neuronal and glial lines, this putative CRE displays binding for the mammalian chromatin organizing protein CTCF. Therefore, we investigated this region for potential allele-specific effects (**Figure 4.3**).

4.3.4 Common variation at rs1990620 affects binding of CTCF at the TMEM106B locus

Given the observed eQTL effect at *TMEM106B* in multiple tissues, and the CTCF binding site within this potential CRE, we next sought evidence for allele-specific binding of CTCF to our candidate regulatory region. To do this, we analyzed ENCODE CTCF ChIP-seq and DNase digital genomic footprinting (DGF) experiments. CTCF ChIP-seq was performed on 99 ENCODE cell lines, 95 of which show a CTCF peak at the putative CRE. We were able to confirm 20 of these

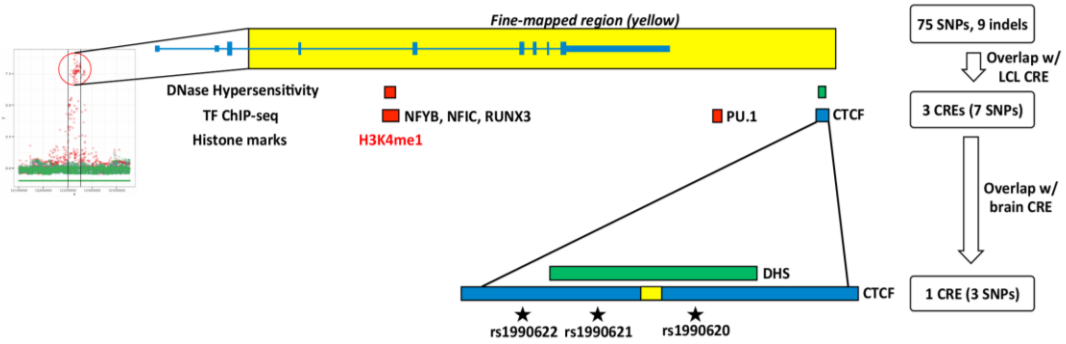


Figure 4.3. Prioritization of *cis*-regulatory elements (CREs) harboring candidate functional variants.
The 84 variants from the eQTL fine-mapping (left) were prioritized based on overlap with predicted CREs in LCLs (red boxes and text), neuronal and glial cell lines (green), or all three (blue), based on ENCODE and NIH Roadmap Epigenome project data (see flow chart at right of image). This analysis yields 7 SNPs in 3 potential CREs as candidate causal variants. Only one CRE – an intergenic CTCF binding site (CTCF motif represented as yellow rectangle) – is predicted to be active in brain tissue; this CTCF-binding CRE contains three SNPs in complete LD, including the GWAS sentinel SNP, rs1990622.

cell lines as *TMEM106B* haplotype heterozygotes by examining reads containing rs1990620, the SNP closest to the CTCF core motif (48bp from core motif) and covered by the most reads (**Figure 4.4A and Table S4.3)**. By analyzing the reads covering rs1990620, we found significant enrichment of CTCF binding to the risk-associated A allele (*P*=0.043, **Figure 4.4B**). In addition, we identified 6 cell lines heterozygous at rs1990620 that were interrogated by DNase DGF. In these lines, we found that the chromosome bearing the risk A allele was significantly more sensitive to DNase cleavage (*P*<0.001, **Figure 4.4B**), consistent with an open chromatin state and potential regulatory activity (Maurano et al., 2015).

We corroborated these data with *in vitro* investigations of CTCF binding by electromobility shift assays (EMSA). Utilizing a competition EMSA approach, we found that the risk allele of rs1990620 was indeed more effective at shifting a protein complex in extracts from HEK293 cells (**Figure 4.4C**), LCLs, (**Figure 4.4D**), and human brain (**Figures 4.4E and S4.3**). Addition of CTCF antibody resulted in both a supershift and disappearance of the original EMSA band, demonstrating that the complex preferentially binding to the A allele of rs1990620 contains CTCF (**Figure 4.4F**). Thus, ChIP-seq-based investigation of differential CTCF binding by allele, coupled with gel shift

assays, together indicate that common variation at a single SNP, rs1990620, may underlie haplotype-specific effects on CTCF recruitment.

4.3.5 Long-range interactions involving TMEM106B demonstrate haplotype-specific effects

Rapidly emerging evidence suggests that CTCF plays a major role in the shaping of the three-dimensional architecture of the mammalian genome (Merkenschlager and Nora, 2016; Ong and Corces, 2014). In particular, CTCF has been reported to contribute to the formation of topologically associated domains (TADs), which may be central to enhancer-promoter interactions and insulator function (Dixon et al., 2012; Dowen et al., 2014; Guo et al., 2015; Lupianez et al., 2015; Nora et al., 2012). Indeed, interrogation of high resolution *in situ* Hi-C data from LCLs (Rao et al., 2014) suggests that the CTCF binding region containing rs1990620 is located within a small ~250kb TAD (sub-TAD) that is part of a larger ~1Mb TAD, and is involved in multiple long-range looping interactions (**Figure S4.4)**. Moreover, one of the major interactions occurs between this region and the *TMEM106B* promoter, implicating a role for this CTCF site in *TMEM106B* regulation (**Figure 4.5A).** Interestingly, there are also strong Hi-C interactions between the promoter and a predicted enhancer of *TMEM106B* (Andersson et al., 2014) located ~13kb downstream of the CTCF site. This enhancer also lies within the sub-TAD and harbors no disease-associated genetic variants. Given the observed allele-specific effects on CTCF recruitment, we hypothesized that rs1990620 may influence *TMEM106B* expression and, by extension, neurodegenerative disease risk, through differential effects on CTCF-mediated interactions between distal regulatory elements (Tang et al., 2015) (**Figure 4.5B**).

To test this, we adapted a recently-developed variation of chromosome conformation capture (Capture-C) (Hughes et al., 2014) to agnostically capture all interactions involving the CTCF-binding region, as well as the *TMEM106B* promoter. Specifically, we coupled 3C library preparation with a probe capture step to enrich for interactions involving our two regions of interest (**Table S4.4**). Importantly, we designed our capture probes not to overlap SNPs (thus, giving

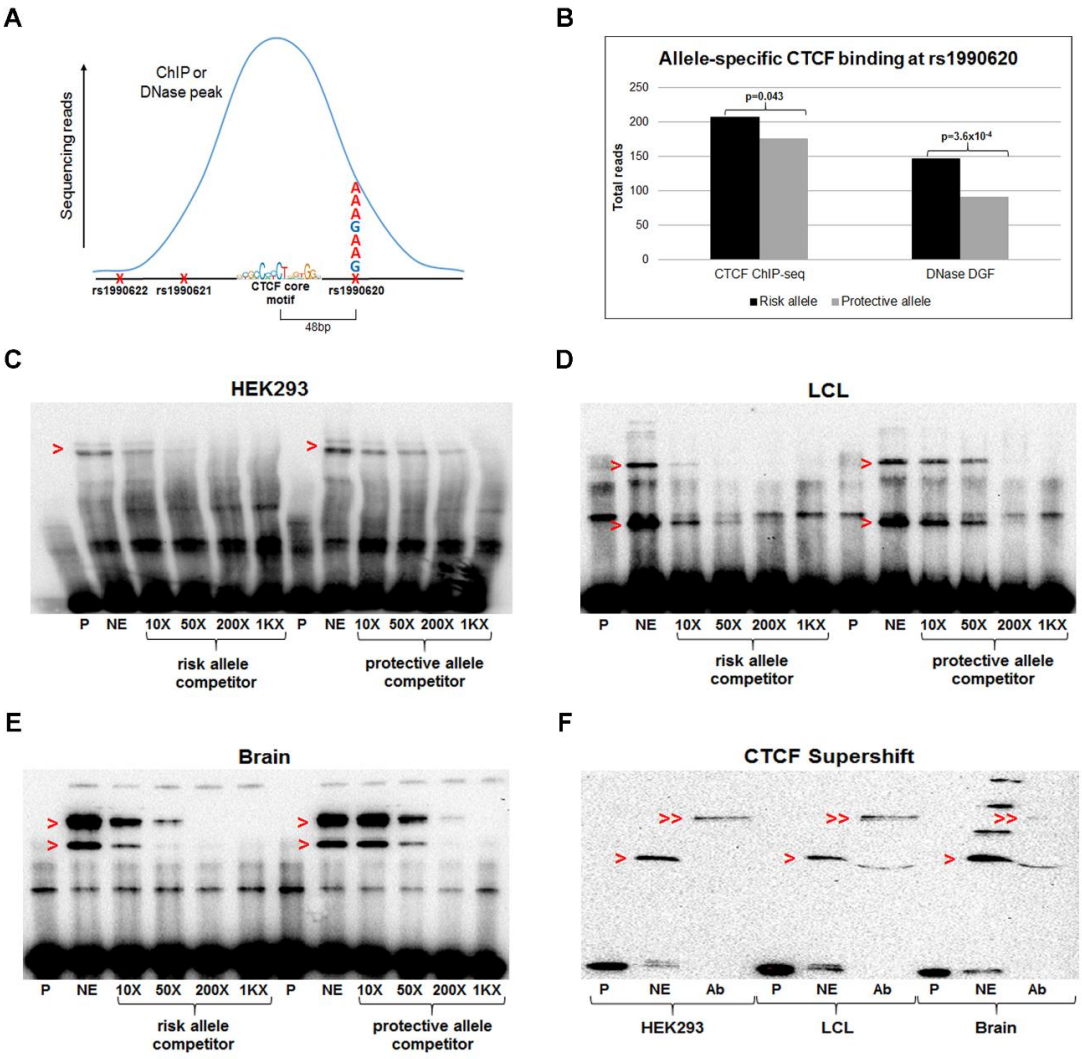


Figure 4.4. The risk allele of rs1990620 preferentially recruits CTCF in human tissues.
(A) Schematic of approach for determining allelic bias in CTCF ChIP-seq and DNase digital genomic footprinting (DGF) experiments. The rs1990620 SNP (48bp from the CTCF core motif) was analyzed for number of reads containing the risk vs. protective allele in samples showing a CTCF ChIP-seq or DNase DGF peak at this region.
(B) The risk allele of rs1990620 increases CTCF binding and DNase hypersensitivity at this region, based on data from 20 and 6 cell types heterozygous at this locus, respectively (**Table S4.3**).
(C-E) A 5' biotinylated probe (P) containing the rs1990620 risk allele was incubated with nuclear extract (NE) from HEK293s **(C)**, LCLs **(D)**, and human brain **(E)**. In each extract, the shifted probe/protein complex (red arrowheads) was more efficiently competed with an unlabeled competitor oligonucleotide (at 10X, 50X, 200X, or 1000X (1KX) the concentration of the labeled probe) containing the risk allele vs. the protective allele, indicating preferential binding of a nuclear factor to the risk allele.
(F) A 3' biotinylated probe (P) containing the rs1990620 risk allele is shifted by addition of nuclear extract (NE) from HEK293s, LCLs and brain (red arrowheads). Addition of an anti-CTCF antibody (Ab) results in disappearance of the shift and a supershift (double arrowheads), indicating that the complex binding to the probe contains CTCF.

probes equal opportunity to bind to either allele), while also localizing to regions within 60bp of one or more marker SNPs (thus allowing for analysis of captured interactions in a haplotype-specific manner). We performed our Capture-C experiments in three different cell lines – two different LCLs and the T cell leukemia-derived Jurkat cell line (the *TMEM106B* eQTL effect has also been reported in T cells (Peters et al., 2016; Raj et al., 2014)) – with all three lines heterozygous for the *TMEM106B* haplotype.

When analyzing all long-range (≥2kb) interactions mapping to chromosome 7, we found that significant interactions (based on an FDR threshold of 1% using fourSig (Williams et al., 2014)) were largely centered around the ~1Mb TAD containing *TMEM106B* (**Figure S4.5**). When restricting the application of fourSig to regions *within* the TAD (thus increasing the background model), all statistically significant interactions mapped to the sub-TAD (**Figure S4.6**). Thus, our data confirm the hierarchical nature of the chromatin architecture previously reported at this locus by *in situ* Hi-C. Importantly, all pair-wise interactions between the five sub-TAD CTCF sites, including the *TMEM106B* promoter, and the enhancer, were statistically significant in every sample, under both analyses (**Figure S4.6**), with the outermost CTCF sites in convergent orientation and delineating the boundaries of a topological domain.

To obtain a finer-grained understanding of the most meaningful interactions at our locus, we further restricted the fourSig analysis to the ~250kb sub-TAD, which further increased the background threshold for significance. Under these conditions, significant interactions between the *TMEM106B* promoter, the rs1990620-containing CTCF site, and the enhancer emerged (**Figures 4.5C and 4.5D**). These results implicate the CTCF site and the enhancer as potential key regulators of *TMEM106B*.

Recent studies suggest that genes involved in CTCF-associated long-range interactions tend to be more transcriptionally active than genes not involved in such interactions (Rao et al., 2014; Tang et al., 2015). Therefore, we asked whether the number of long-range chromatin looping interactions involving the risk haplotype, which preferentially binds CTCF and expresses *TMEM106B* at higher levels, is significantly higher than the interactions involving the protective

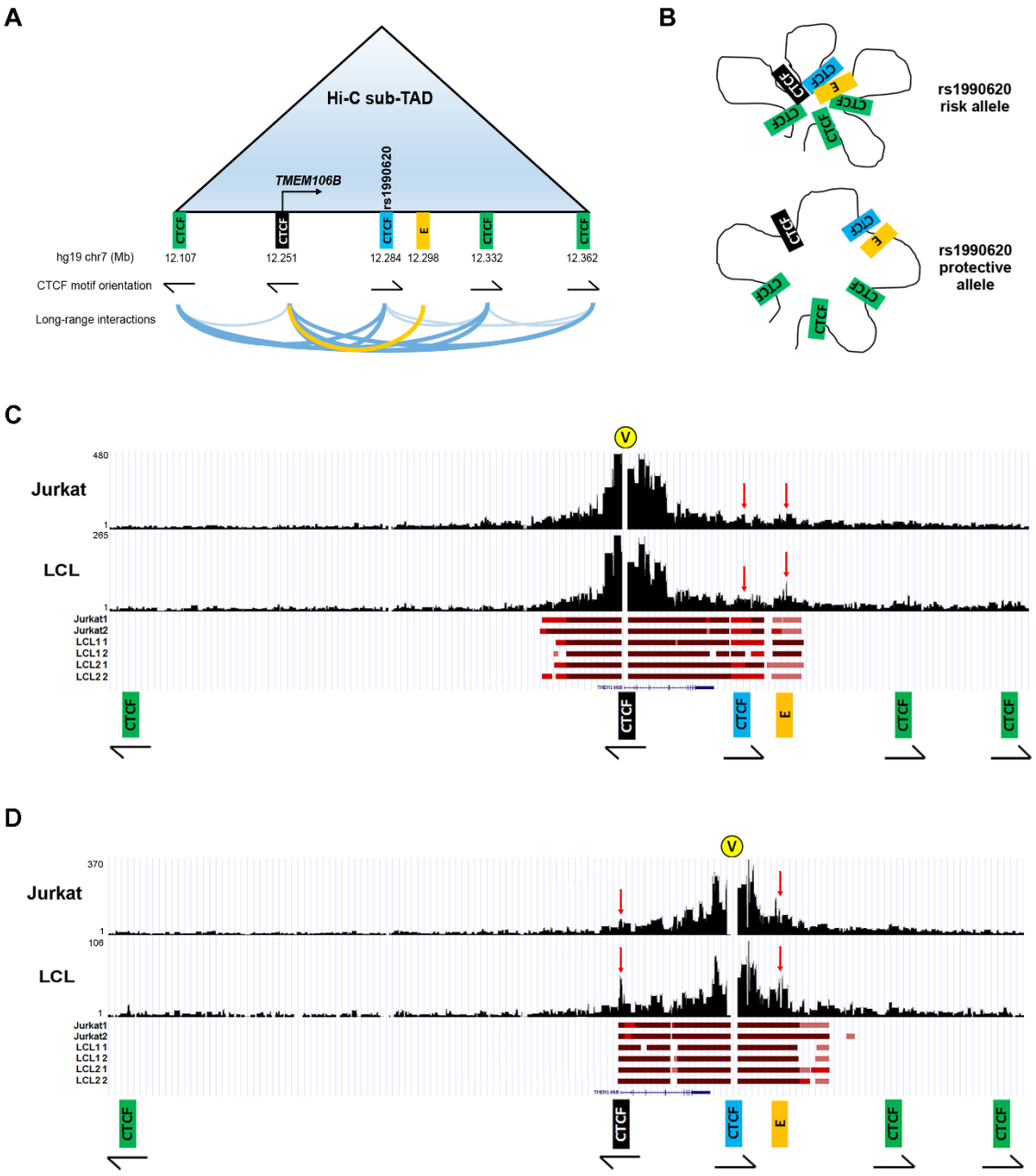

A
Hi-C sub-TAD
TMEM106B
rs1990620
CTCF
E
hg19 chr7 (Mb)
12.107
12.251
12.284
12.298
12.332
12.362
CTCF motif orientation
Long-range interactions
B
rs1990620
risk allele
rs1990620
protective
allele
C
Jurkat
LCL
480
265
Jurkat1
Jurkat2
LCL1 1
LCL1 2
LCL2 1
LCL2 2
V
D
Jurkat
LCL
370
106
E
Chr7
TAD
sub-TAD
0.60
0.55
0.50
0.45
0.40
***
*
**
n.s.
Jurkat
LCL #1
LCL #2
Risk
Protective

Figure 4.5. Haplotype-specific long-range chromatin interactions at the *TMEM106B* locus.
(A) Schematic representation of the *TMEM106B* sub-TAD and interactions among distal regulatory elements, based on published LCL Hi-C data. The black CTCF site is located at the *TMEM106B* promoter; the blue CTCF site contains rs1990620; the gold rectangle labeled "E" represents a transcriptionally active enhancer. Note that the CTCF motifs present at the sub-TAD boundaries (12.107 and 12.362) follow the convergent orientation (arrows indicate direction and strand) most commonly reported for interacting CTCF sites. Darker lines in bottom of panel indicate interactions between CTCF sites in convergent orientation.
(B) Model illustrating how allele-specific CTCF binding at rs1990620 might affect sub-TAD structure and long-range interactions at this locus, with more contact among distal regulatory elements bearing the risk-associated haplotype.
(C,D) Capture-C experimental data for representative Jurkat and LCL samples, with raw read coverage shown on the y-axis for interactions captured by probes for (**C**) the *TMEM106B* promoter and **(D)** the rs1990620-containing CTCF site. Significant interactions within the sub-TAD for each cell line and replicate (3 cell lines with 2 replicates each) are indicated with red bars, below the coverage plots, with darker shades of red indicating higher confidence interactions. Yellow circles marked "V" indicate viewpoints (capture sites). Red arrows indicate interactions between the promoter, the rs1990620 CTCF site, and enhancer.
(E) Allelic bias in long-range interactions involving the *TMEM106B* promoter across all of chromosome 7 (left), the 1Mb TAD (middle), and the 250kb sub-TAD (right) containing *TMEM106B*. Read count proportions from capture experiments containing either the risk (blue) or protective (orange) allele of a marker SNP are shown; in each case, more interactions with the *TMEM106B* promoter involve the risk haplotype. *p<0.01; **p<0.001; ***p<0.0001; n.s.=non-significant.

haplotype, in these heterozygous cell lines (**Figure 4.5A**). In all three cell lines, we observed significantly more interactions captured with the promoter probes occurring on the risk haplotype, with consistent effects whether we analyzed all interactions mapping to chromosome 7, or restricted the analysis to interactions within the TAD or sub-TAD (**Figure 4.5E**). When interactions were captured with probes targeting the rs1990620-containing CTCF binding site, fewer reads were obtained, and no apparent difference in raw reads involving the risk vs. protective haplotype were detected.

Given that various sources of technical bias (*e.g.* capture bias, alignment bias, bias in duplicate removal) can influence allele-specific high-throughput sequencing analyses, we next compared the number of long-range interactions involving each haplotype ("true" interactions) against the number of reads from each haplotype aligning to regions directly adjacent to the bait regions ("false" interactions,). We assumed that regions directly adjacent to the bait regions are subject to artefactual ligations due simply to chromosomal proximity (Cairns et al., 2016). While some true functional interactions may be lost in this way (creating false negatives), this approach

let us test for false-positive differences in interactions between the two haplotypes by capturing biases in read alignment or other technical steps of the Capture-C protocol.

After adjustment for technical bias, we still observed significant enrichment of promoter-captured interactions on the risk haplotype in two out of three cell lines (P=1.98x10^{-2} for Jurkats and P=5.78x10^{-3} for LCL Line 2 for significant deviation from expected proportions). Moreover, the same two cell lines demonstrated a significant enrichment of CTCF-associated interactions on the risk haplotype as well, whether probes captured interactions adjacent to the rs1990620 SNP (P=3.28x10^{-3} for Jurkats and P=4.37x10^{-3} for LCL Line 2) or the rs1990621 SNP (P<1.00x10^{-6} for Jurkats and P=2.38x10^{-2} for LCL Line 2, see **Figures 4.3** and **4.4A** for SNP locations).

Taken together, these data suggest that SNP-specific effects on CTCF recruitment may alter the genomic architecture at the *TMEM106B* locus, with ensuing alterations in gene expression.

4.3.6 Common genetic variants associated with neurodegenerative diseases are enriched in brain CTCF binding sites

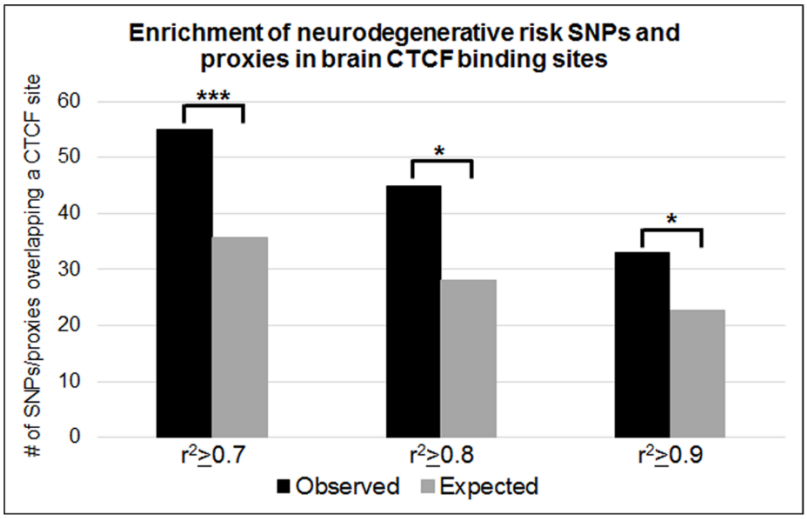
CTCF is emerging as a master regulator of mammalian gene expression through its widespread influences on genomic architecture (Merkenschlager and Nora, 2016; Ong and Corces, 2014; Won et al., 2016). Having uncovered an allele-specific effect on the expression of *TMEM106B,* a frontotemporal dementia-associated genetic risk factor, likely mediated by CTCF, we hypothesized that CTCF-mediated effects may play a more general role in conferring risk for neurodegeneration.

To test this, we identified all published SNPs associated at a genome-wide statistical significance level with risk for four major neurodegenerative diseases (FTLD, AD, PD, and ALS). 200 neurodegenerative disease risk SNPs were identified (**Table S4.5**) from the GWAS Catalog (Welter et al., 2014), and this set of SNPs, as well as their LD proxies, were investigated for the extent of overlap with brain CTCF binding sites identified by ChIP-seq (ENCODE Project Consortium et al., 2012).

Seven human brain-relevant CTCF ChIP-seq datasets were identified from ENCODE project data (ENCODE Project Consortium et al., 2012). Across the combined set of all seven

datasets, using the Genomic Regulatory Elements and GWAS Overlap algorithm (GREGOR (Schmidt et al., 2015)), we found a highly-significant ~1.5-fold enrichment of neurodegenerative disease SNPs and their LD proxies overlapping CTCF binding sites (P<0.0001 at an LD threshold of $r^2 \geq 0.7$, **Figure 4.6A**). Evaluating each ChIP-seq dataset individually, CTCF binding sites found in brain-derived microvascular endothelial cells (1.6-fold enrichment, P<0.001), cerebellar astrocytes (1.4-fold enrichment, P=0.006), and neuroblastoma cells (1.5-fold enrichment, P=0.008) were most highly enriched for the presence of neurodegenerative risk SNPs, whereas CTCF binding sites found in other brain-derived cell types, including choroid plexus epithelial cells, showed no significant enrichment (**Figure 4.6B**). Our results are thus compatible with a wider role for CTCF-mediated effects in modulating risk for neurodegeneration.

A



B

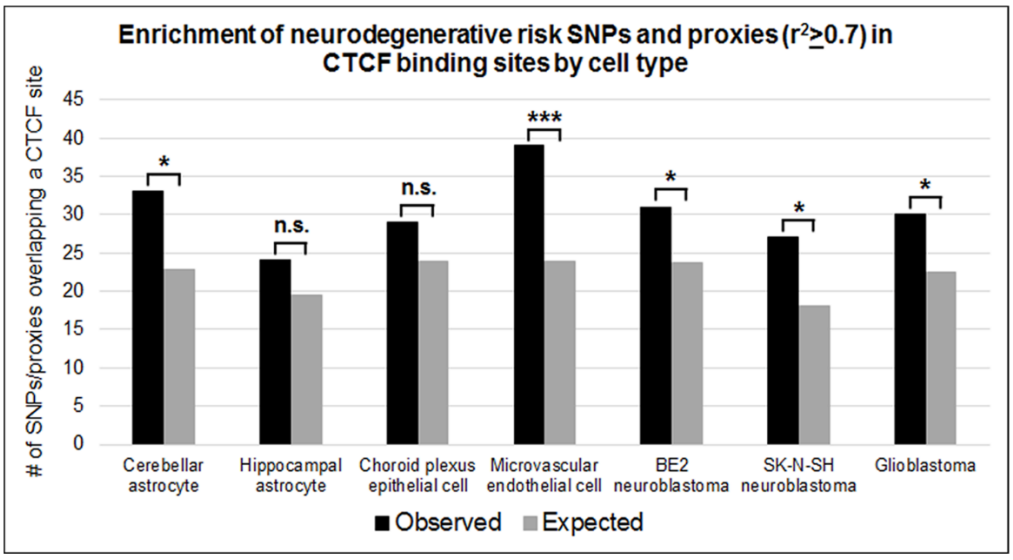


Figure 4.6. Neurodegenerative disease risk SNPs are enriched in brain CTCF binding sites.

We determined the overlap between all SNPs identified by GWAS to confer risk for four different neurodegenerative diseases, as well as their proxies (at LD thresholds of $r^2 \geq 0.7$, ≥ 0.8 or ≥ 0.9), with brain CTCF binding sites. We then compared this "observed" overlap with an "expected" overlap with brain CTCF binding sites, based on 500 control SNPs matched to each GWAS SNP for minor allele frequency, number of LD proxies at each threshold, and distance to nearest gene, using the GREGOR algorithm. Statistically significant enrichments of neurodegenerative disease risk SNPs at CTCF binding sites were seen in the combined data sets at each LD threshold for proxy determination **(A)**, and in 5 out of 7 brain-relevant cell types, shown at an LD threshold of $r^2 \geq 0.7$ **(B)**. *p<0.05; **p<0.005; ***p<0.0005; n.s.=non-significant.

4.4 DISCUSSION

Here, we functionally dissect a locus first linked to the human neurodegenerative disease FTLD by GWAS. Through a combination of data-mining and bench-based experimental studies of human-derived tissues, we demonstrate that common variants linked to FTLD by GWAS associate with haplotype-specific expression of *TMEM106B* in multiple tissues including human brain; that this effect may depend on haplotype-specific effects on recruitment of CTCF, with corresponding haplotype-specific effects on long-range chromatin interactions; and that incremental changes in *TMEM106B* expression have effects on cell toxicity (**Figure 4.7**). Thus, we provide molecular characterization of the *TMEM106B* locus, elucidating a mechanism by which a causal variant at this locus may exert downstream effects on risk for FTLD.

We note that the effects we describe are incremental, rather than all-or-none. Similar incremental effects have been reported for many cases of allele-specific expression (Dimas et al., 2009; GTEx Consortium, 2015; Tewhey et al., 2016), including allele-specific expression differences associated with neuropsychiatric disease variants (Sekar et al., 2016; Soldner et al., 2016). Conceptually, such an incremental effect is not surprising for common genetic variants that confer only slightly increased odds for developing a disease, yet may still shed light on important disease mechanisms.

Aspects of the work presented here may be more broadly applicable to common variant effects on risk for many neurodegenerative diseases or, even more broadly, for many common variant-trait associations. For example, here, genotype at the causal noncoding variant rs1990620 alters FTLD risk through an incremental effect on *TMEM106B* expression. The enrichment of disease-associated variants in predicted *cis*-regulatory regions (Maurano et al., 2012; Schaub et al., 2012) and the overlap between these variants and variants associated with gene expression levels (eQTLs) (Nicolae et al., 2010; Zhu et al., 2016) suggest that many common variants identified by GWAS may act by modulating gene expression. Moreover, most post-GWAS functional studies have identified putative causal variants that are thought to affect disease risk by altering the expression of nearby or distant genes (Adrianto et al., 2011; Bauer et al., 2013; Claussnitzer et al.,

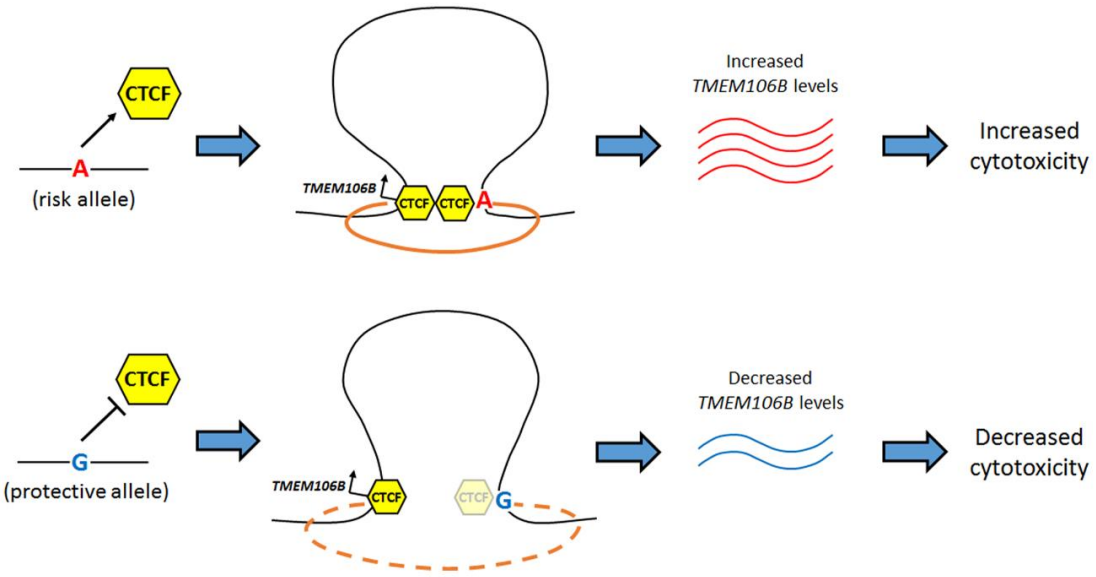


Figure 4.7. Working model of molecular mechanism underlying 7p21 association with dementia.
The risk-associated allele of the causal variant (rs1990620) preferentially recruits CTCF, resulting in haplotype-specific effects on long-range chromatin interactions, with downstream effects of increased *TMEM106B* expression. Increased *TMEM106B* expression leads to increased cytotoxicity and corresponding risk for neurodegeneration.

2015; Harismendy et al., 2011; Huang et al., 2014; Musunuru et al., 2010). Indeed, we find here that SNPs associated with risk for neurodegeneration by GWAS are enriched in brain CTCF binding sites, suggesting that allele-specific modulation of gene expression programs influenced by CTCF may underlie additional risk factors for other neurodegenerative diseases.

We note that the degree of molecular precision provided here is largely absent in the characterization of neurodegenerative disease loci first discovered by GWAS. Yet this level of mechanistic detail illuminating the genetic regulation and biological function of GWAS-derived loci is certainly needed if we are to translate the thousands of "leads" obtained in this way into potential avenues for therapeutic interventions. In this context, the strategy illustrated here of prioritization of variants based on the wealth of newly available genomic data, followed by targeted investigation in cell culture systems, may be more broadly applicable to the study of common variants associated with other human diseases.

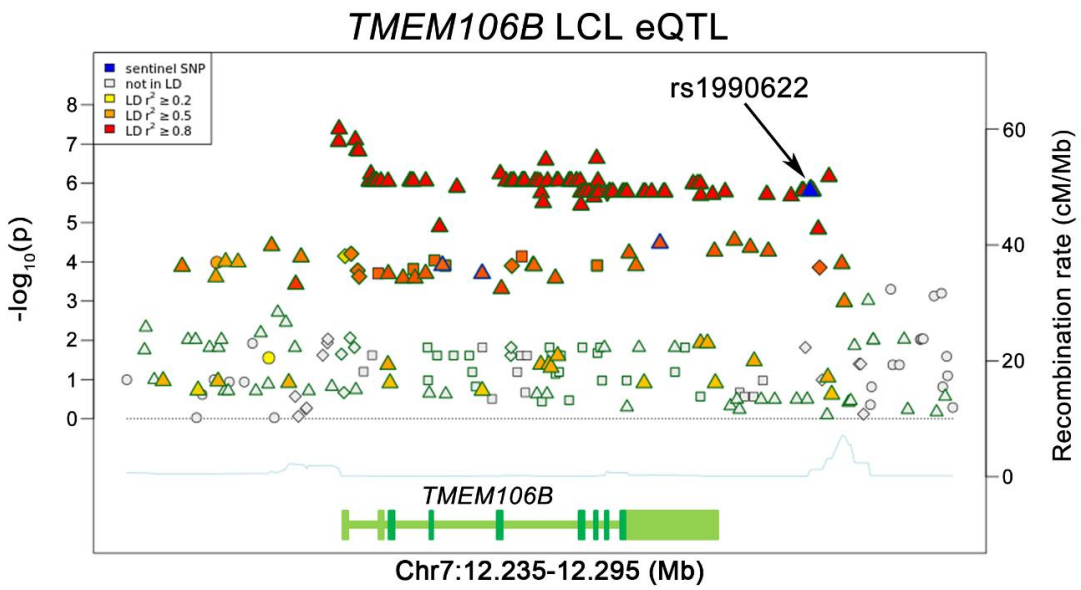


Figure S4.1. Multiple linked 7p21 variants are associated with *TMEM106B* expression in LCLs.

All GTEx SNPs tested for association with *TMEM106B* levels in LCLs within a 60kb window are shown (GRCh37/hg19 coordinates). SNPs are color coded according to linkage disequilibrium with rs1990622, the GWAS sentinel SNP (shaded in blue and indicated with arrow). X-axis indicates genomic position, with *TMEM106B* gene structure indicated as a reference; y-axes indicate p-value of association with *TMEM106B* levels (left) and recombination rate (right, indicated by blue line under eQTL plot).

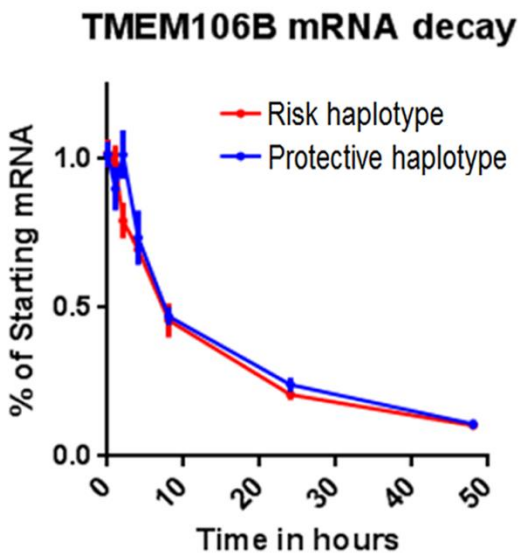


Figure S4.2. *TMEM106B* genotype does not affect mRNA stability.
Lymphoblastoid cell lines homozygous for the risk (n=3) and protective (n=3) *TMEM106B* haplotypes were treated with Actinomycin D to inhibit transcription. RNA was isolated at 0h, 1h, 2h, 4h, 8h, and 24h after treatment, and *TMEM106B* was quantified by RT-qPCR. Data were analyzed with a two-way ANOVA.

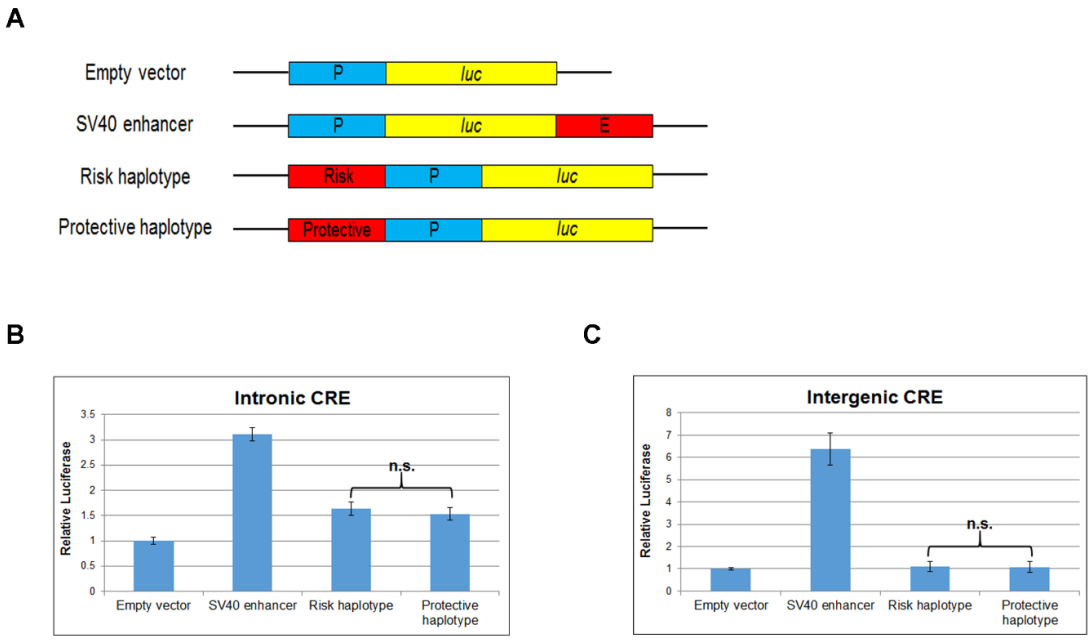


Figure S4.3. Intronic and intergenic candidate *cis*-regulatory elements (CREs) do not harbor functional variants.
(A) Schematic of luciferase reporter constructs used to test candidate CREs for enhancer activity. Empty vector construct contains no enhancer (negative control); SV40 enhancer construct serves as a positive control; test constructs contain putative CREs with either the risk or protective SNP alleles cloned upstream of the SV40 promoter. P=SV40 promoter; *luc*=luciferase; E=SV40 enhancer.
(B,C) The intronic CRE displays weak ~1.5-fold enhancer activity with both SNP haplotypes ($p<0.001$), whereas neither SNP haplotype of the intergenic CRE displays any activity. n.s.=non-significant.

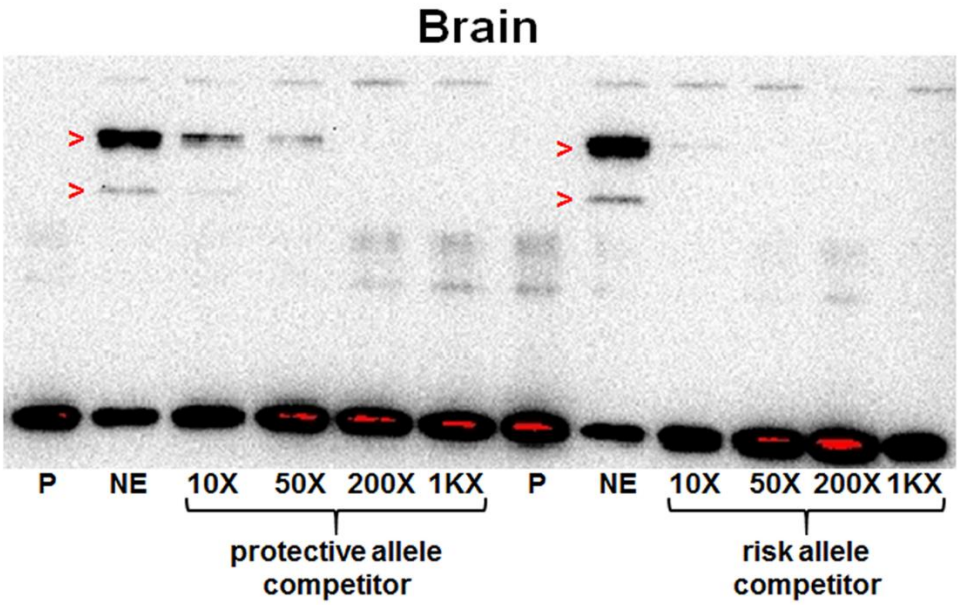


Figure S4.4. The risk allele of rs1990620 preferentially recruits a nuclear factor in brain nuclear extract.
A 5' biotinylated probe containing the rs1990620 protective allele was incubated with human brain nuclear extract, resulting in a similar shift (red arrowheads) to that seen with the rs1990620 risk allele (**Fig. 4.4E**). The shifted complex was better competed with excess amounts of unlabeled oligo containing the risk allele of rs1990620 than that containing the protective allele.

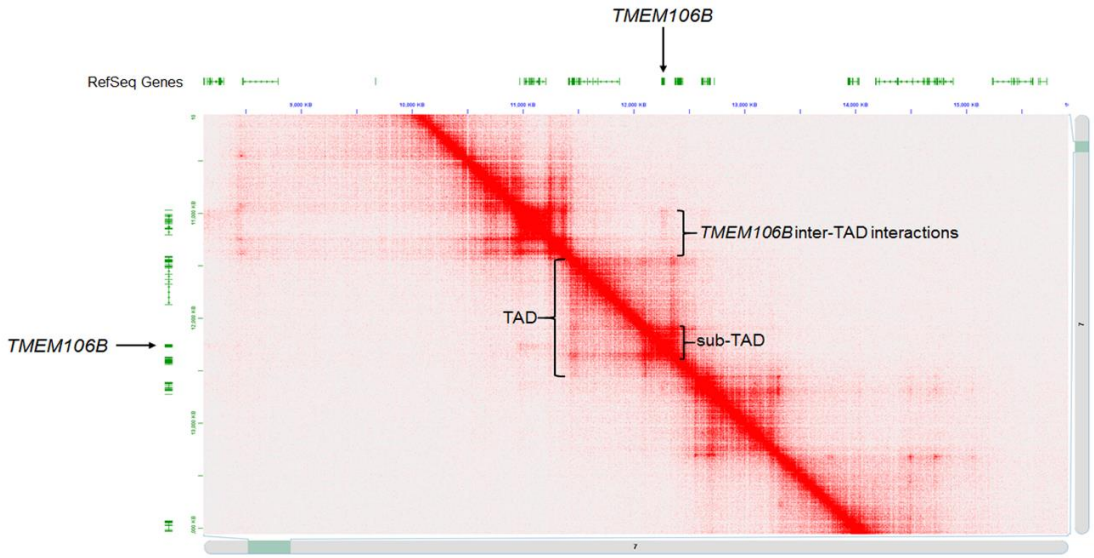
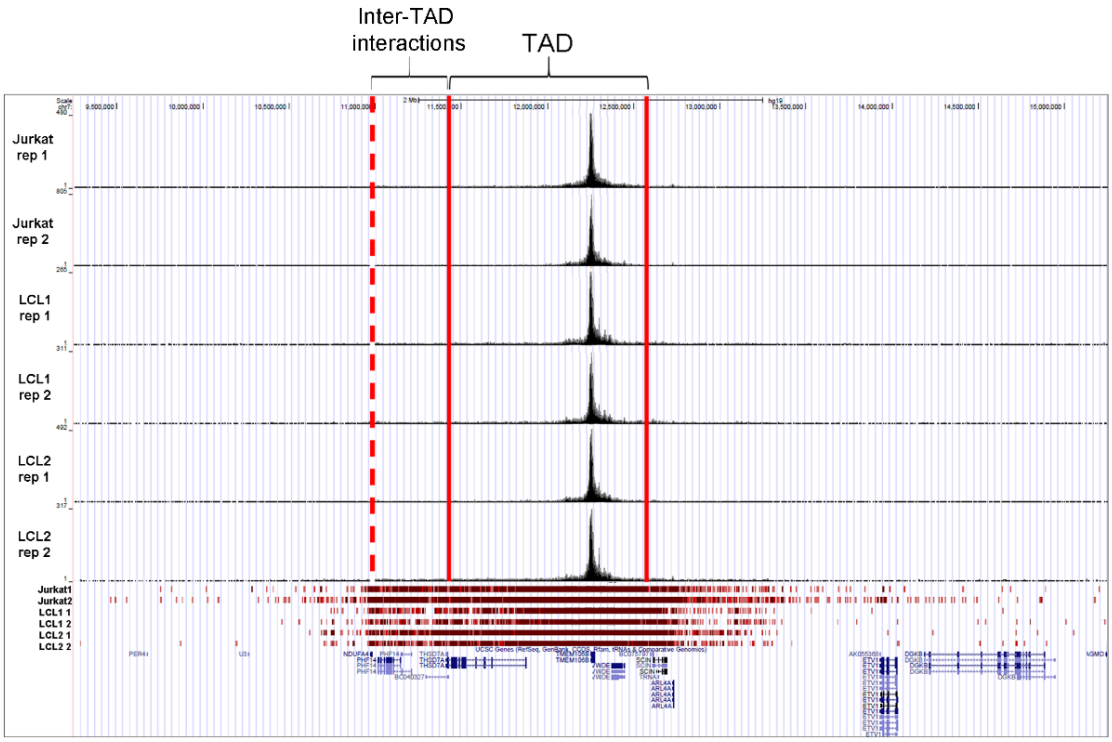


Figure S4.5. Lymphoblastoid cell line *in situ* Hi-C data reveals the chromatin architecture at the *TMEM106B* locus.
Hi-C heatmap for chromosome 7p21 displays a ~250kb topologically associating domain (sub-TAD) within a larger ~1Mb TAD, containing *TMEM106B* and no other genes. *TMEM106B* also interacts with genomic regions extending several hundred kilobases upstream of the TAD (inter-TAD interactions).


Inter-TAD interactions
TAD
Jurkat rep 1
Jurkat rep 2
LCL1 rep 1
LCL1 rep 2
LCL2 rep 1
LCL2 rep 2
Jurkat1
Jurkat2
LCL1 1
LCL1 2
LCL2 1
LCL2 2

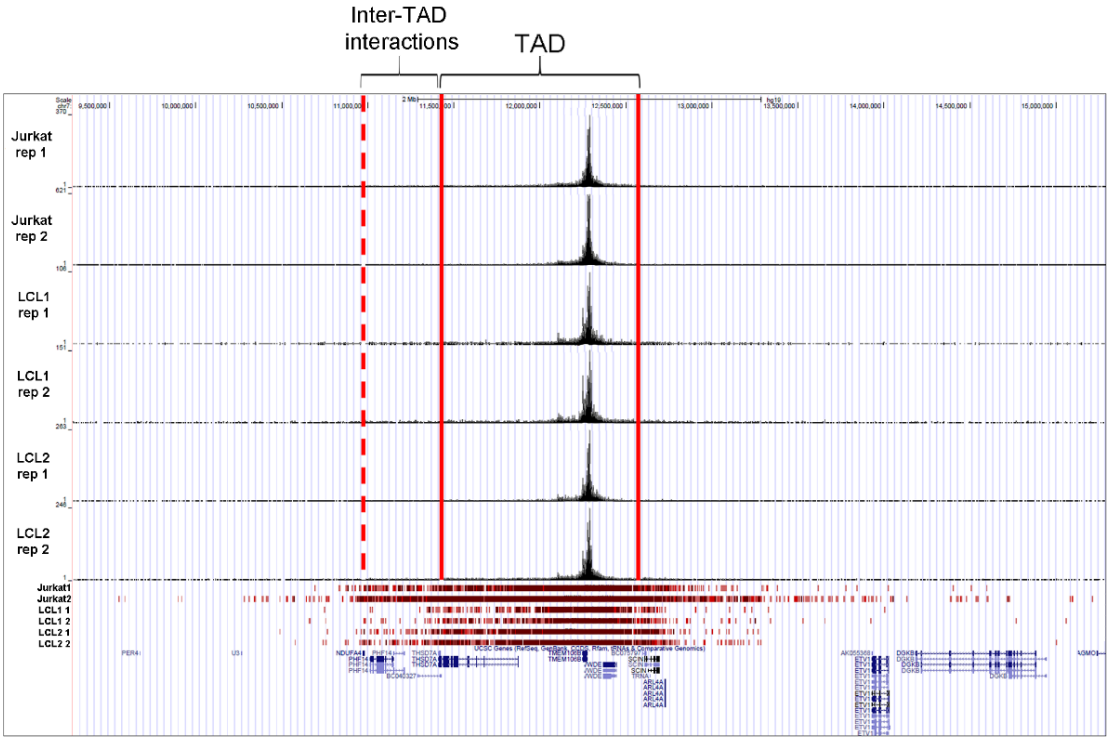
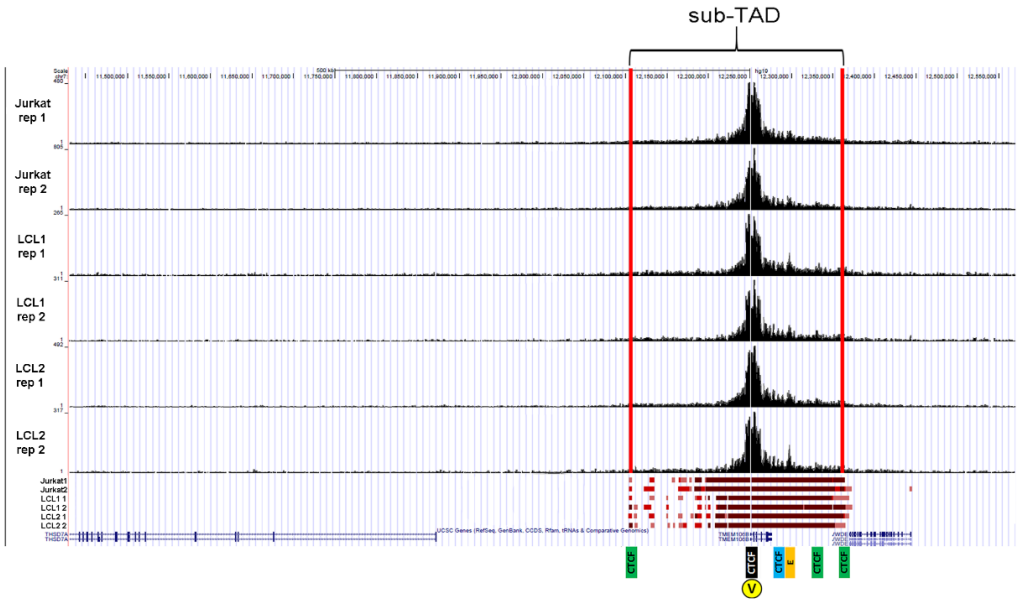

Inter-TAD interactions
TAD
Jurkat rep 1
Jurkat rep 2
LCL1 rep 1
LCL1 rep 2
LCL2 rep 1
LCL2 rep 2
Jurkat1
Jurkat2
LCL1 1
LCL1 2
LCL2 1
LCL2 2

Figure S4.6. Capture-C recapitulates the predicted chromatin architecture at the *TMEM106B* locus.
When analyzing all long-range interactions mapping to chromosome 7 from the **(top)** *TMEM106B* promoter or **(bottom)** CTCF site viewpoint, statistically significant interactions are limited mostly to coordinates agreeing with the Hi-C TAD (solid red lines) and upstream inter-TAD interactions (dashed red line indicates endpoint of interactions observed in the Hi-C data, see **Figure S4.5**). The top of each panel shows raw read coverage for each sample and replicate (3 cell lines, with 2 replicates each), and the bottom of each panel shows statistically significant interactions (red bars indicate statistical significance, with darker shades of red indicating higher confidence interactions). Data are viewed in the UCSC Genome Browser, with read counts on the y-axis of each track.

A



B

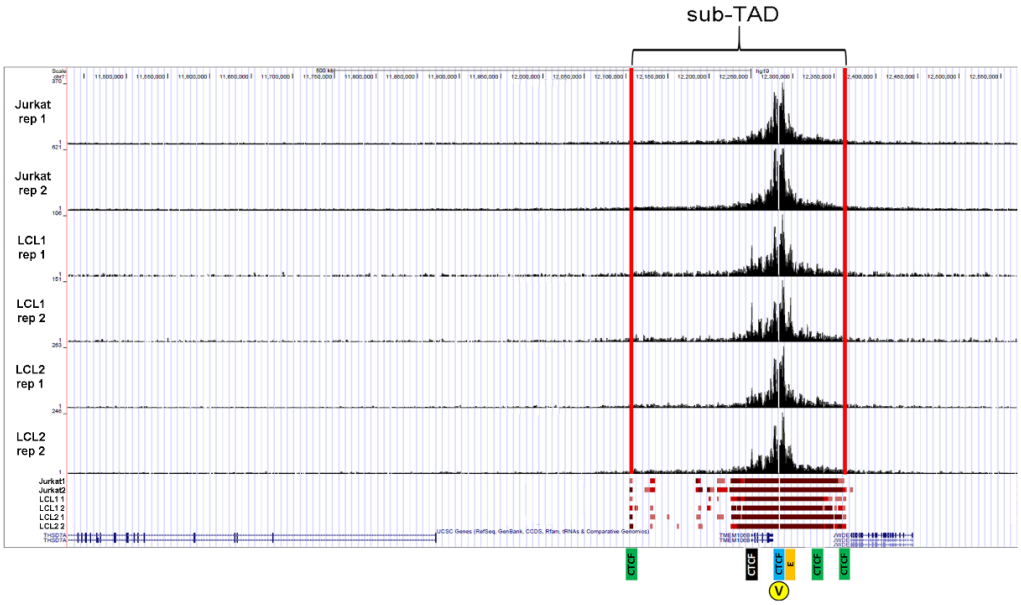


Figure S4.7. Capture-C identifies key interactions between *cis*-regulatory elements at the *TMEM106B* locus.

When analyzing only the interactions mapping to the ~1Mb TAD, statistically significant interactions from the **(A)** promoter and **(B)** CTCF site viewpoint map entirely to the Hi-C sub-TAD coordinates (solid red lines). Key interacting regions from **Figure 4.5A** are indicated below each browser snapshot. Yellow circles marked with a "V" indicate the Capture-C viewpoint. Data are viewed in the UCSC Genome Browser, with read counts on the y-axis of each track. Red arrows highlight increase in read counts, corresponding to interactions with the rs1990620-containing CTCF site and nearby enhancer (**A**) and the *TMEM106B* promoter (**B**).

Putative CRE	rsid	hg19	DNase HS		TF binding		Histone marks		LD with rs1990622
			LCL	Brain	LCL	Brain	LCL	Brain	
intronic CRE	rs1435527	12262571	Yes	No	NFIC RUNX3 NFYB	No	H3K4me1	No	0.96
intronic CRE	rs1435525	12262717	Yes	No	NFIC RUNX3 NFYB	No	H3K4me1	No	0.96
intronic CRE	rs1435524	12262801	Yes	No	NFIC RUNX3 NFYB	No	H3K4me1	No	0.96
intergenic CRE	rs1548884	12279761	No	No	PU.1	No	No	No	0.98
CTCF site	rs1990622	12283787	No	Yes	CTCF	CTCF	No	No	N/A
CTCF site	rs1990621	12283873	No	Yes	CTCF	CTCF	No	No	0.99
CTCF site	rs1990620	12284008	No	Yes	CTCF	CTCF	No	No	1

Table S4.1. Functional annotations of tested top eQTL SNPs variants.
The 7 SNPs overlapping either DNase hypersensitivity (HS), transcription factor (TF) binding, or active histone marks in lymphoblastoid cell lines (LCLs) are listed by location. CTCF binding at the region containing the GWAS sentinel SNP, rs1990622, and two other linked SNPs, is seen in brain cell types as well as LCLs.

rsid	hg19	LD with rs1990622 (r^2)
rs6460902	12255511	0.93
rs35982495	12258665	no info
rs3823612	12258755	0.96
rs11509880	12261911	0.67
rs3800847	12262242	0.96
rs11368032	12262968	no info
rs5882346	12263183	0.94
rs11974335	12263378	0.94
rs11974304	12263437	0.94
rs10950392	12263538	0.96
rs10950393	12263546	0.96
rs10950394	12263587	0.96
rs11509153	12263800	0.96
rs3800843	12264297	0.96
rs10950395	12264467	0.96
rs10950396	12264666	0.96
rs10950397	12264737	0.98
rs10950398	12264871	0.95
rs76854159	12265059	0.95
rs150465020	12265192	0.96
rs6948844	12265848	0.96
rs6966757	12265952	0.96
rs6966915	12265988	0.98
rs7804433	12266706	0.95
rs7804234	12266814	0.96
rs7804736	12266867	0.98
rs4721057	12267221	0.96

rs4721058	12267256	0.96
rs4721059	12267495	0.97
rs4721060	12267552	0.97
rs4721061	12267559	0.97
rs7792410	12267614	0.97
rs7809700	12267734	0.97
rs13229988	12267837	0.98
rs34046032	12268082	no info
rs28549831	12268117	0.98
rs13230513	12268243	0.98
rs12667950	12268468	0.98
rs5011431	12268551	0.98
rs5011434	12268717	0.96
rs5011436	12268758	0.95
rs5011437	12268792	0.97
rs5011438	12268798	0.97
rs5011439	12268811	0.97
rs3839691	12269256	0.97
rs3173615	12269417	0.98
rs13237715	12269575	0.98
rs13237518	12269593	0.97
rs12699332	12269762	0.98
rs12699333	12269804	0.98
rs12668625	12269817	0.98
rs3815535	12270519	0.98
rs3214372	12270565	no info
rs2302635	12270625	0.98
rs2302634	12270770	0.98

rs2302633	12270815	0.98
rs7808568	12271252	0.96
rs12669919	12271997	0.98
rs1042949	12272116	0.98
rs3800841	12272568	0.97
rs34252347	12272862	no info
rs1054168	12273413	0.98
rs1054169	12273496	0.98
rs35337387	12274413	no info
rs1468804	12275508	0.96
rs5882347	12275624	no info
rs1468803	12275675	0.96
rs1060700	12275818	0.96
rs1468802	12276011	0.96
rs1468801	12276045	0.96
rs2356065	12276885	0.98
rs7797705	12277773	0.98
rs10281425	12280730	0.98
rs6460905	12283227	1
rs6460906	12283276	1
rs7791726	12283329	1
rs12699338	12285140	0.98

Table S4.2. List of untested top eQTL variants.
The top eQTL SNPs that had no functional annotations in LCLs are listed by location. The linkage disequilibrium (LD) of each variant with rs1990622 according to HaploReg v4.1 is indicated. Variants without r^2 values were not annotated in HaploReg v4.1, but have nearly identical minor allele frequencies as all other variants, suggesting that they are also in strong LD. hg19=genomic coordinates from the GRCh37/hg19 human reference genome build.

Experiment	Cell lines
CTCF ChIP-seq	BE2_C, Caco-2, GM06990, GM12864, GM12872, GM12873, GM12874, GM12891, GM12892, GM19238, H1-hESC, HBMEC, HCFA, HEK293, HepG2, HRPEpiC, HUVEC, HVMF, NHDF, RPTEC
DNase digital genomic footprinting	AoAf, HCFaa, NHDF-neo, HMVEC-dBl-Ad, H7-hESC, HMVEC-dLy-Neo

Table S4.3. Cell lines used for CTCF ChIP-seq and DNase digital genomic footprinting analyses.

All cell lines were confirmed to be heterozygous at rs1990620 by analyzing reads containing rs1990620. Major and minor allele rs1990620 reads were enumerated and pooled across samples for each experiment. Deviation from an expected ratio of 50:50 was tested using a two-tailed binomial sign test. Cell lines are listed by their ENCODE identifiers, or spelled out if no identifiers exist.

Probe	Sequence
Promoter probe 1	GAGTTGCTGTGTCGCCTCTAATGAGGCCCAGCCAGGGAACAC TCGGCTTCGGCCCAAGCC
Promoter probe 2	ACCTCTGGGTTGCCTCTAGGCCCTCACACCTTGAGCGCCAGG TGGCCCTCTTCCTTTTGC
Promoter probe 3	CTAGGCCCTCACACCTTGAGCGCCAGGTGGCCCTCTTCCTTTT GCTGTTGATGAATGTTC
Promoter probe 4	TTGAGCGCCAGGTGGCCCTCTTCCTTTTGCTGTTGATGAATGT TCTTGCCGTGGTGCCGG
CTCF site probe 1	CAATGGAGAACAGAGCTGCAGTACAACTATAGCATTTTGAG GGGCTGGGAGAACGGGATA
CTCF site probe 2	TACAGTACTACAATGGAGAACAGAGCTGCAGTACAACTATA GCATTTTGAGGGGCTGGGA
CTCF site probe 3	ATGAAGAGTCTAAGCAGAACTTTCCAGTCATTGTACTACAAT GGGTTGCAGGCGTGCCAG
CTCF site probe 4	TTTCCAGTCATTGTACTACAATGGGTTGCAGGCGTGCCAGCC CTCTGGTGGCCACATAAG
CTCF site probe 5	ATGGGTTGCAGGCGTGCCAGCCCTCTGGTGGCCACATAAGTC CTGGAGTGGCTGGTCATC
CTCF site probe 6	GCCCTCTGGTGGCCACATAAGTCCTGGAGTGGCTGGTCATCT CAGACTGGGAGCTTCCTC

Table S4.4. Probe sequences used for Capture-C.
Four and six 60bp biotinylated probes were used to capture interactions involving the *TMEM106B* promoter and rs1990620-containing CTCF site, respectively.

SNP	REPORTED GENE(S)	DISEASE/TRAIT(S)	P-VALUE	ODDS RATIO or BETA
rs10122902	intergenic	Amyotrophic lateral sclerosis	5.00E-11	NR
rs10260404	DPP6	Amyotrophic lateral sclerosis	5.00E-08	1.30
rs10498633	SLC24A4, RIN3	Alzheimer's disease (late onset)	6.00E-09	1.10
rs10513789	MCCC1, LAMP3	Parkinson's disease	3.00E-10	1.25
rs10792832	PICALM	Alzheimer's disease (late onset)	9.00E-26	1.15
rs10797576	SIPA1L2	Parkinson's disease	5.00E-10	1.13
rs10838725	CELF1	Alzheimer's disease (late onset)	1.00E-08	1.08
rs10948363	CD2AP	Alzheimer's disease (late onset)	5.00E-11	1.10
rs11023139	RRAS2, PSMA1, SPON1	Alzheimer's disease (cognitive decline)	7.00E-11	0.31
rs11060180	CCDC62	Parkinson's disease	6.00E-12	1.11
rs11061269	GPR133	Amyotrophic lateral sclerosis (sporadic)	8.00E-10	3.78
rs11136000	CLU	Alzheimer's disease	9.00E-10	1.16
rs11154851	PDE7B	Alzheimer's disease (cognitive decline)	1.00E-08	0.25
rs11158026	GCH1	Parkinson's disease	6.00E-11	1.11
rs11218343	SORL1	Alzheimer's disease (late onset)	1.00E-14	1.30
rs11225434	MMP-1	Matrix metalloproteinase levels	9.00E-29	0.39
rs11226373	MMP-1, MMP-3	Matrix metalloproteinase levels	2.00E-18	0.44
rs11248051	GAK	Parkinson's disease	3.00E-09	1.46
rs11248060	GAK, DGKQ	Parkinson's disease	3.00E-12	1.21
rs112724034	PGAM5P1, MAN2A1	Alzheimer's disease (cognitive decline)	9.00E-13	0.31
rs115102486	CLMN	Alzheimer's disease (cognitive decline)	2.00E-08	0.31
rs11711441	MCCC1, LAMP3	Parkinson's disease	8.00E-12	1.19
rs11724635	BST1	Parkinson's disease	9.00E-18	1.13
rs11744876	CTNND2	Amyotrophic lateral sclerosis (sporadic)	3.00E-08	NR
rs11754661	MTHFD1L	Alzheimer's disease (late onset)	2.00E-10	2.10
rs11767557	EPHA1	Alzheimer's disease (late onset)	6.00E-10	1.11

rs11771145	EPHA1	Alzheimer's disease (late onset)	1.00E-13	1.11
rs117896735	INPP5F	Parkinson's disease	4.00E-13	1.62
rs117964204	CACNA1G	Alzheimer's disease (cognitive decline)	1.00E-09	0.28
rs11931074	SNCA	Parkinson's disease	7.00E-17	1.37
rs12075	FCER1A, OR10J1, DARC, OR10J5	Obesity-related traits	1.00E-21	0.10
rs12185268	MAPT	Parkinson's disease	3.00E-14	1.30
rs12203592	IRF4	Progressive supranuclear palsy	6.00E-15	1.48
rs12456492	RIT2	Parkinson's disease	8.00E-12	1.11
rs12608932	UNC13A	Amyotrophic lateral sclerosis	3.00E-14	1.25
rs12636651	CCR3, CCR2	Obesity-related traits	7.00E-09	0.05
rs12637471	MCCC1	Parkinson's disease	2.00E-21	1.19
rs12726330	GBA	Parkinson's disease	5.00E-08	1.71
rs12817488	CCDC62, HIP1R	Parkinson's disease	3.00E-13	1.14
rs12989701	BIN1	Alzheimer's disease (late onset)	3.00E-10	1.23
rs13048019	SOD1	Amyotrophic lateral sclerosis	3.00E-08	2.02
rs1318862	FAT3	Subcortical brain region volumes	6.00E-09	26.86
rs1411478	STX6	Progressive supranuclear palsy	4.00E-11	1.27
rs14235	BCKDK, STX1B	Parkinson's disease	2.00E-12	1.10
rs1442190	LRRK2	Parkinson's disease	2.00E-27	3.72
rs1474055	STK39	Parkinson's disease	1.00E-20	1.21
rs1476679	ZCWPW1	Alzheimer's disease (late onset)	6.00E-10	1.10
rs148763909	SAP30L	Alzheimer's disease (cognitive decline)	1.00E-08	0.15
rs1491942	LRRK2	Parkinson's disease	6.00E-15	1.17
rs1541160	KIFAP3	Amyotrophic lateral sclerosis	2.00E-08	0.58
rs1555399	TMEM229B	Parkinson's disease	7.00E-14	1.11
rs1562990	MS4A	Alzheimer's disease	4.00E-11	1.14
rs157580	APOE	Alzheimer's disease	8.00E-89	1.69
rs157582	APOE	Psychosis and Alzheimer's disease	9.00E-52	2.30

rs1582763	MS4, MS4A6A, MS4A4A, MS4A6E	Alzheimer's disease in APOE e4-carriers	2.00E-09	1.15
rs1605070	IQCF5	Amyotrophic lateral sclerosis (sporadic)	2.00E-09	3.95
rs1630500	GBA	Parkinson's disease	2.00E-08	1.75
rs17125944	FERMT2	Alzheimer's disease (late onset)	8.00E-09	1.14
rs17314229	FRMD4A	Alzheimer's disease	1.00E-10	1.68
rs17393344	MYO16	Alzheimer's disease (cognitive decline)	2.00E-08	0.26
rs17577094	MAPT	Parkinson's disease	8.00E-09	1.56
rs17649553	MAPT	Parkinson's disease	2.00E-48	1.30
rs1768208	MOBP	Progressive supranuclear palsy	5.00E-17	1.37
rs17817600	PICALM	Alzheimer's disease	2.00E-08	1.33
rs1795240	FMO3, FMO6P	Lentiform nucleus volume	5.00E-08	112.27
rs190982	MEF2C	Alzheimer's disease (late onset)	3.00E-08	1.08
rs1925690	ZNF292	MRI atrophy measures	3.00E-08	0.00
rs1936246	GAPDHP15, RBBP4P4	Alzheimer's disease (APOE e4 interaction)	2.00E-13	1.04
rs1990622	TMEM106B	FTLD-TDP	1.00E-11	1.60
rs199347	GPNMB	Parkinson's disease	1.00E-12	1.11
rs1994090	LRRK2	Parkinson's disease	3.00E-08	1.39
rs199515	WNT3	Parkinson's disease	3.00E-17	1.32
rs199533	NSF	Parkinson's disease	1.00E-14	1.28
rs1997111	intergenic	Cerebrospinal T-tau levels	1.00E-08	NR
rs2075650	APOE, TOMM40	Alzheimer's disease	2.00E-157	2.53
rs2102808	STK39	Parkinson's disease	4.00E-10	1.18
rs2228467	CCBP2	Cerebrospinal fluid levels of Alzheimer's disease-related proteins	4.00E-18	NR
rs2279590	CLU	Alzheimer's disease	6.00E-10	1.22
rs2338971	HUSEYO	Parkinson's disease	4.00E-10	1.35
rs2373115	GAB2	Alzheimer's disease (late onset)	1.00E-10	4.06
rs2414739	VPS13C	Parkinson's disease	1.00E-11	1.11

rs242557	MAPT	Progressive supranuclear palsy	9.00E-18	1.43
rs2446581	FRMD4A	Alzheimer's disease	1.00E-10	1.68
rs2494250	FCER1A, OR10J3	Select biomarker traits	1.00E-14	NR
rs2685056	ALCAM	Amyotrophic lateral sclerosis (sporadic)	4.00E-08	2.58
rs2707466	WNT16	Cortical thickness	2.00E-10	0.11
rs2718058	NME8	Alzheimer's disease (late onset)	5.00E-09	1.08
rs2736990	SNCA	Parkinson's disease	2.00E-16	1.23
rs2814707	IFNK, MOBKL2B, C9orf72	Amyotrophic lateral sclerosis	7.00E-09	1.22
rs28834970	PTK2B	Alzheimer's disease (late onset)	7.00E-14	1.10
rs2899472	CYP19A1	Cerebrospinal AB1-42 levels	2.00E-09	NR
rs2942168	MAPT	Parkinson's disease	1.00E-28	1.27
rs3026968	CADM3	Inflammatory biomarkers	9.00E-14	NR
rs329648	MIR4697	Parkinson's disease	1.00E-11	1.11
rs34311866	TMEM175, GAK, DGKQ	Parkinson's disease	1.00E-43	1.27
rs34372695	SYT11	Parkinson's disease	4.00E-12	1.47
rs34517613	SALM1	Amyotrophic lateral sclerosis (sporadic)	9.00E-09	1.20
rs34972666	TGM6	Alzheimer's disease (cognitive decline)	3.00E-08	0.23
rs35349669	INPP5D	Alzheimer's disease (late onset)	3.00E-08	1.08
rs356182	SNCA	Parkinson's disease	4.00E-73	1.32
rs356219	SNCA	Parkinson's disease	6.00E-65	1.29
rs356220	SNCA	Parkinson's disease	8.00E-35	1.38
rs35749011	GBA, SYT11	Parkinson's disease	1.00E-29	1.82
rs3764650	ABCA7	Alzheimer's disease	5.00E-17	1.23
rs3818361	CR1	Alzheimer's disease	4.00E-14	1.18
rs3849942	IFNK, MOBKL2B, C9orf72	Amyotrophic lateral sclerosis	9.00E-11	2.16
rs3849943	C9orf72	C9orf72	6.00E-10	1.22
rs3851179	PICALM	PICALM	1.00E-09	1.16

rs3865444	CD33	CD33	2.00E-09	1.10
rs393152	MAPT, C17orf69, KIAA1267, LOC644246	MAPT, C17orf69, KIAA1267, LOC644246	2.00E-16	1.30
rs4128725	OR10J1	OR10J1	4.00E-12	NR
rs4129267	IL6R	IL6R	2.00E-57	NR
rs4147929	ABCA7	ABCA7	1.00E-15	1.15
rs429358	APOE	APOE	5.00E-14	NR
rs4420638	APOE, TOMM40, APOC1	Alzheimer's disease	8.00E-149	3.45
rs4538475	BST1	Parkinson's disease	3.00E-09	1.24
rs4663105	BIN1	Alzheimer's disease in APOE e4- carriers	2.00E-12	1.19
rs4676049	intergenic	Alzheimer's disease (late onset)	4.00E-08	1.76
rs4700060	ANKRD55	Alzheimer's disease (cognitive decline)	1.00E-08	0.21
rs471470	PICALM	Alzheimer's disease in APOE e4+ carriers	3.00E-08	1.16
rs4796217	CCL4L2	Protein quantitative trait loci	4.00E-21	NR
rs4866650	intergenic	Left inferior lateral ventricle volume (Cerebrospinal fluid biomarker status interaction)	1.00E-09	6.15
rs4906844	LOC100128714	Cortical thickness	1.00E-08	0.07
rs4938933	MS4A4A	Alzheimer's disease (late onset)	8.00E-12	1.12
rs495366	MMP	Matrix metalloproteinase levels	6.00E-34	0.44
rs4968782	CYB561, ACE	Cerebrospinal fluid levels of Alzheimer's disease-related proteins	4.00E-12	NR
rs514716	GLIS3	Alzheimer's disease biomarkers	3.00E-09	0.07
rs519113	APOE, CEACAM16, BCL3, PVRL2, TOMM40, PPP1R37	Alzheimer's disease (late onset)	5.00E-39	2.09
rs536841	PICALM	Alzheimer's disease	3.00E-09	1.16
rs56131196	APOC1	Alzheimer's disease biomarkers	4.00E-12	NR
rs561655	PICALM	Alzheimer's disease (late onset)	7.00E-11	1.15
rs569214	CLU	Alzheimer's disease	4.00E-08	1.14

rs573521	MMP3, MMP12, WTAPP1, MMP13, DCUN1D5	Cerebrospinal fluid levels of Alzheimer's disease-related proteins	2.00E-44	NR
rs58370486	BZW2, TSPAN13	Alzheimer's disease (cognitive decline)	6.00E-11	0.36
rs59007384	TOMM40	Alzheimer's disease biomarkers	7.00E-09	NR
rs610932	MS4A4E, MS4A6A	Alzheimer's disease	2.00E-14	1.11
rs61144803	CRADD	Alzheimer's disease (cognitive decline)	5.00E-08	0.16
rs61812598	IL6R, TDRD10, SHE, UBE2Q1, ADAR	Cerebrospinal fluid levels of Alzheimer's disease-related proteins	6.00E-63	NR
rs62256378	SUCLG2	Cerebrospinal AB1-42 levels in Alzheimer's disease dementia	3.00E-12	0.71
rs62341097	GALNT7	Neuritic plaque	6.00E-09	1.15
rs6430538	ACMSD, TMEM163	Parkinson's disease	9.00E-20	1.14
rs6532194	SNCA	Parkinson's disease	5.00E-11	1.29
rs6547705	CD8B	Progressive supranuclear palsy	1.00E-08	1.28
rs6599388	GAK	Parkinson's disease	4.00E-12	1.16
rs6599389	GAK	Parkinson's disease	4.00E-08	1.31
rs6656401	CR1	Alzheimer's disease (late onset)	6.00E-24	1.18
rs6701713	CR1	Alzheimer's disease (late onset)	5.00E-10	1.16
rs6703183	CAMK1G	Amyotrophic lateral sclerosis	3.00E-08	1.31
rs6733839	BIN1	Alzheimer's disease (late onset)	7.00E-44	1.22
rs6738962	CTNNA2	Alzheimer's disease (cognitive decline)	1.00E-08	0.18
rs679515	CR1	Alzheimer's disease in APOE e4+ carriers	4.00E-09	1.22
rs6808835	CCRL2, CCR5, CCR2, LTF, CCR3	Cerebrospinal fluid levels of Alzheimer's disease-related proteins	2.00E-13	NR
rs6812193	SCARB2, FAM47E	Parkinson's disease	3.00E-11	1.10
rs6857	PVRL2	Dementia and core Alzheimer's disease neuropathologic changes	2.00E-62	1.61
rs6859	APOE, PVRL2, TOMM40	Alzheimer's disease	6.00E-14	NR
rs6887649	FTMT	Left inferior lateral ventricle volume (Cerebrospinal fluid biomarker status interaction)	2.00E-08	5.68

rs6922617	TREM, NCR2	Alzheimer's disease biomarkers	4.00E-08	0.09
rs7081208	FRMD4A	Alzheimer's disease	1.00E-10	1.68
rs7117082	LOC100128095, OPCML	Amyotrophic lateral sclerosis (sporadic)	8.00E-09	2.79
rs7274581	CASS4	Alzheimer's disease (late onset)	3.00E-08	1.14
rs72832584	BCAS3	Alzheimer's disease (cognitive decline)	1.00E-11	0.30
rs744373	BIN1	Alzheimer's disease	3.00E-14	1.17
rs7561528	BIN1	Alzheimer's disease (late onset)	4.00E-14	1.17
rs7571971	EIF2AK3	Progressive supranuclear palsy	4.00E-13	1.33
rs75932628	TREM2	Alzheimer's disease	2.00E-12	2.90
rs76904798	LRRK2	Parkinson's disease	5.00E-14	1.16
rs769449	APOE, TOMM40	Alzheimer's disease biomarkers	2.00E-18	0.09
rs7849530	SPTLC1	Left inferior lateral ventricle volume (Cerebrospinal fluid biomarker status interaction)	3.00E-10	6.40
rs8045064	FLJ45256	Alzheimer's disease (cognitive decline)	4.00E-08	0.21
rs8070723	MAPT	Progressive supranuclear palsy	2.00E-118	5.11
rs8118008	DDRGK1	Parkinson's disease	3.00E-11	1.11
rs8141797	SUSD2, CABIN1, GGT5	Amyotrophic lateral sclerosis	2.00E-09	1.52
rs823118	NUCKS1, RAB7L1	Parkinson's disease	2.00E-16	1.12
rs855913	ZNF746	Amyotrophic lateral sclerosis	4.00E-08	1.08
rs9327881	NUDT12	Amyotrophic lateral sclerosis (sporadic)	5.00E-09	NR
rs9329300	PFKP	Amyotrophic lateral sclerosis (sporadic)	2.00E-09	3.10
rs9331888	CLU	Alzheimer's disease	6.00E-10	1.22
rs9331896	CLU	Alzheimer's disease (late onset)	3.00E-25	1.16
rs9349407	CD2AP	Alzheimer's disease (late onset)	9.00E-09	1.11
rs947211	SLC45A3, PARK16, PM20D1, NUCKS1, RAB7L1, SLC41A1	Parkinson's disease	2.00E-12	1.30
rs9525638	TNFSF11, RANKL	Cortical thickness	4.00E-09	0.09

rs9825420	ITGA9	Amyotrophic lateral sclerosis (sporadic)	3.00E-08	3.03
rs983392	MS4A6A	Alzheimer's disease (late onset)	6.00E-16	1.11
rs9877502	UTS2D, SNAR-I, GEMC1, OSTNIL1-RAP, CCDC50	Alzheimer's disease biomarkers	5.00E-09	0.05
rs997277	DIRC1, COL3A1	Parkinson disease and Lewy body pathology	5.00E-08	1.63
rs9977018	TMPRSS2	Amyotrophic lateral sclerosis (sporadic)	2.00E-09	NR

Table S4.5. SNPs associated with risk for neurodegenerative disease and related phenotypes by GWAS.

For each SNP, the reported gene(s), associated disease/traits, and p-value, odds ratio or beta corresponding to the association are listed. For SNPs associated with multiple traits, the information listed pertains to the trait with which the SNP is most significantly associated (lowest p-value). NR=not reported.

4.5 METHODS

eQTL analyses

The GWAS sentinel SNP, rs1990622 (Van Deerlin et al., 2010), was queried for association with all transcripts genome-wide using The Genotype-Tissue Expression (GTEx) eQTL database (GTEx Consortium, 2015), consisting of 7,051 samples and representing 44 different tissues from 449 healthy donors. GTEx eQTL plots were generated with *SNiPA* (Arnold et al., 2015). Conditional analyses and fine-mapping were performed using HapMap3-imputed genotypes from a published multi-ethnic LCL eQTL study (Stranger et al., 2012), as previously described (Brown et al., 2013). In brief, gene expression data were normalized to the empirical average quantiles across all samples. Subsequently, the distribution of each gene expression trait was transformed to the quantiles of the standard normal distribution, separately within each population. The effects of known and unknown covariates were controlled for by principal component analysis. A *cis*-eQTL scan was performed by regressing the additive effect of each SNP within 1Mb of *TMEM106B* on gene expression by Bayesian regression, as implemented in SNPTEST (Wellcome Trust Case Control Consortium, 2007).

Analysis of LD structure at the *TMEM106B* locus

HapMap Phase I, II and III combined CEU genotype data (International HapMap 3 Consortium et al., 2010) were visualized in HaploView to assign LD blocks (Barrett et al., 2005), filtering out SNPs with MAF<0.001 and requiring that 90% of informative pairwise LD values within a block must represent strong LD. Pairwise LD between variants at the *TMEM106B* locus was determined using both HapMap and 1000 Genomes data (1000 Genomes Project Consortium et al., 2012), visualized either in HaploView or on the HaploReg v4.1 web resource (Ward and Kellis, 2012).

Cell culture

HeLas were cultured in DMEM with 10% FBS, 1% L-glutamine, and 1% penicillin/streptomycin. LCLs were obtained from the Coriell Institute for Medical Research in Camden, NJ, and were cultured in RPMI with 15% FBS, 1% L-glutamine and 1%

penicillin/streptomycin. Jurkats were cultured in RPMI with 10% FBS, 1% L-glutamine and 1% penicillin/streptomycin.

***TMEM106B* overexpression experiments**

TMEM106B expression constructs designed to over-express at 2-fold, 5-fold, and 20-fold expression, as previously described (Chen-Plotkin et al., 2012), were transfected into HeLas with Lipofectamine 2000, as per manufacturer protocols. 48 hours post-transfection, ten brightfield images were taken across three biological replicates for each condition at 100X on a Life Technologies EVOS FL microscope. Image files were assigned random identifiers and a blinded individual counted the number of cells in each image that displayed the vacuolar phenotype, defined by having at least one clear punctate vacuolar structure. This experiment was repeated three times and the results were pooled. To assess cytotoxicity, the same transfection protocol was carried out, but at 48h post-transfection cells were spun down and resuspended in Trypan Blue-containing DMEM culture media, and the proportion of Trypan Blue-positive cells were determined with a hemocytometer. This experiment was also carried out three times. Western blots were performed for all six experiments as described previously (Chen-Plotkin et al., 2012), and the data were pooled together for the quantification shown in **Figure 4.2B**.

Cell line haplotype phasing

In order to confirm all cell lines used for mRNA stability experiments, ASE analysis, and Capture-C as *TMEM106B* haplotype heterozygotes, we first performed TaqMan SNP Genotyping assays, as previously described (Gallagher et al., 2014), for rs1990622 and marker SNPs in strong or complete LD with rs1990622 (rs3807865, r^2=0.9; rs6966915, r^2=1; rs3173615, r^2=1; rs1468803, r^2=1), to confirm heterozygosity. For the cell lines used for the Capture-C experiments, the CTCF binding region was also analyzed by Sanger sequencing to confirm heterozygosity for the three completely linked potential causal variants: rs1990622, rs1990621, and rs1990620. Heterozygosity of the promoter SNP rs4721056 (r^2=0.5 with rs1990622) was confirmed by genotyping as well, but because of the lower LD of this SNP with rs1990622, we PCR amplified the region containing rs4721056 and three SNPs in strong LD with rs1990622 (rs7781670, r^2=0.9;

rs1019309, r^2=0.9; and rs1019307, r^2=0.89) in the three Capture-C cell lines. This amplicon was cloned into the MCS of the pGL3 Promoter vector, and Sanger sequencing of individual clones confirmed that no cell lines had mixed haplotypes comprising these SNPs, thus linking the risk allele of rs4721056 to the risk haplotype in all cell lines.

mRNA stability experiments

Three LCLs homozygous for the *TMEM106B* risk haplotype, and three LCLs homozygous for the protective haplotype, were treated with 1µg/µL Actinomycin D, and RNA was extracted at 0h, 1h, 2h, 4h, 8h and 24h post-treatment. RT-qPCR was performed as previously described (Chen-Plotkin et al., 2012) to quantify *TMEM106B* levels at each time point, normalized to *18S* RNA, which decayed by only ~14% in 24hr. Decay curves for mRNA stability were compared using two-way ANOVAs. This experiment was performed on each of the six cell lines twice, for a total of six biological replicates for each haplotype.

Epigenomic prioritization of candidate *cis*-regulatory elements

We used the UCSC Genome Browser (GRCh37/hg19) (Hinrichs et al., 2016) to visualize ENCODE DNase hypersensitivity, transcription factor ChIP-seq, and the H3K4me1, H3K4me3, and H3K27ac histone mark tracks in the GM12878 LCL line (ENCODE Project Consortium et al., 2012). We also visualized these histone marks in all LCLs included in the NIH Roadmap Epigenome Project (Roadmap Epigenomics Consortium et al., 2015) using the WashU Epigenome Browser (Zhou and Wang, 2012). First, we determined which of the top eQTL SNPs from the fine-mapping overlapped a *cis*-regulatory element (CRE) predicted to be active in LCLs, based on the presence of *at least one* of the features mentioned above. This permissive filtering process yielded only three candidate CREs with seven candidate causal variants. Two putative CREs displaying either DNase hypersensitivity, transcription factor binding, and/or H3K4me1 were tested for enhancer activity in LCL reporter assays. The third CRE contains three candidate causal variants and displays CTCF binding in LCLs, neuronal and glial cell lines, and was further investigated for allele-specific effects.

Reporter assays

To test the two putative LCL CREs for enhancer activity, we cloned each region with either the risk or protective haplotype SNP alleles, based on 1000 Genomes Phase I (1000 Genomes Project Consortium et al., 2010) haplotype information, into the upstream multiple cloning site (MCS) of the pGL3 Promoter luciferase reporter vector (Promega Cat. #E1761). We transfected $2.5x10^6$ LCLs using program Y-001 and nucleofection solution V on the Lonza Nucleofector 2b, using the pGL3 Basic vector (with no enhancer) (Promega Cat. #E1751) as a negative control, and the pGL3 Control vector (with an SV40 enhancer) (Promega Cat. #E1741) as a positive control (**Figure S4.3B**). Three or four biological replicates were included for each construct in each experiment, and four independent experiments were performed for each candidate CRE. Cell lysates were isolated 24h post-transfection for luciferase readout using the Promega Dual-Luciferase Reporter Assay System (Promega Cat. #E1910), as described previously (Chen-Plotkin et al., 2012). We used two-tailed t-tests to test for statistically significant differences in reporter activity.

ENCODE data-mining for CTCF binding and DNase hypersensitivity

We downloaded the BAM read alignment files for all CTCF ChIP-seq experiments that showed a CTCF peak at the region containing rs1990620, as well as for the DNase digital genomic footprinting experiments performed in cell lines that show a DNase-seq peak at this region (https://www.encodeproject.org/, **Table S4.3)**. We analyzed raw sequencing reads containing rs1990620 to identify cell lines heterozygous at this locus. We summed risk allele reads across all cell types, as well as protective allele reads across all cell types, assessing for deviation from a 50:50 proportion using a two-tailed binomial sign test.

Electrophoretic mobility shift assays

Nuclear extract was obtained from healthy human occipital cortex, LCLs, and HEK293s using the Thermo Fisher Scientific Nuclear and Cytoplasmic Extraction Reagents kit (Cat. #78833). A 61bp biotinylated DNA probe containing the risk or protective allele of rs1990620 at position 31 and 30bp of genomic sequence on either side was incubated with extract and competed with excess amounts of unlabeled oligo containing either the risk or protective allele of rs1990620. 2µL

of anti-CTCF antibody (EMD Millipore Cat. #07-729) was added to the reaction to test for supershifts. EMSAs were performed in accordance with the Thermo Fisher Scientific LightShift Chemiluminescent EMSA kit (Cat. #20148).

Capture-C

Capture-C was performed similarly to previous reports (Hughes et al., 2014). Briefly, 3C libraries were generated by fixing $10x10^6$ cells with formaldehyde, followed by *DpnII* digestion and ligation. Phenol/chloroform-extracted DNA was sonicated to produce 200-300bp fragments, and sequencing libraries were prepared with the NEBNext DNA Library Prep Master Mix Set (Illumina Cat. #E6040). 10µg of each capture library underwent multiplexed PCR with unique index oligos. Hybridization with 60bp biotinylated capture probes (**Table S4.4**) was performed with the SeqCap EZ Hybridization and Wash Kit (Roche Cat. #05634261001). Briefly, 3C libraries were air-dried with heat in a thermocycler, and resuspended with hybridization reagents. 2µL (3pmoles) of pooled capture probes for each bait region were added to the resuspended libraries, and incubated for 72h in a thermocycler at 47 degrees C. After isolation of captured material using streptavidin beads and PCR, an additional 24h capture was performed. The capture probes, ordered from Integrated DNA Technologies, flank *DpnII* cut sites that are proximal to marker SNPs in the *TMEM106B* promoter region and CTCF site, designed for captured ligation products to contain marker SNPs distinguishing between haplotypes. Promoter capture experiments used rs4721056 to distinguish haplotypes, and CTCF site capture experiments used rs1990620 and rs1990621 to distinguish haplotypes (see “Cell line haplotype phasing” section). Samples were pooled and sequenced on one lane of an Illumina HiSeq 2500 with 125bp paired end reads, yielding ~230 million read pairs. Two LCLs and Jurkat cells were captured at both regions, with two technical replicates for each capture, performed by different individuals in tandem.

Capture-C data analyses

Quality control was performed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and pre-processing and read alignment were performed as previously described (Davies et al., 2016). Briefly, the paired-end

reads were reconstructed into single reads using FLASH, *in silico*-digested using the DpnII2E.pl script (https://github.com/telenius/captureC/releases), and mapped with Bowtie1(Langmead et al., 2009). The aligned reads were then analyzed with the CCanalyser3.pl script (https://github.com/telenius/captureC/releases). Interactions were tested for significance with fourSig (Williams et al., 2014), using default parameters and a window size of five. First, we tested interactions for significance using all reads mapping to chromosome 7. Then we restricted our analyses to interactions within the TAD and sub-TAD to test for significant interactions *within* these regions, respectively.

To determine if the long-range interactions captured by the hybridization probes show allelic bias, we first estimated the technical (non-biological) bias for each SNP, which may reflect capture bias, mapping bias, or other sources of technical bias. We estimated technical bias using two independent approaches. For the first approach, we enumerated the risk vs. protective allele reads containing each marker SNP directly *adjacent* to the probe sequences (*i.e.* mapping to the 5kb window containing the probe sequences or the immediately adjacent 5kb windows at each capture site) for each experiment. In the second approach, we enumerated the risk vs. protective allele reads containing each marker SNP mapping *at a great distance* from *TMEM106B* (*i.e.* outside of the *TMEM106B* TAD, as defined by the LCL Hi-C heatmap (Rao et al., 2014)). In each case, we assumed that the interactions too close or too far from *TMEM106B* do not represent functional interactions. For the adjacent interactions, ligations may occur artefactually due to chromosomal proximity (Cairns et al., 2016). For the interactions outside the *TMEM106B* TAD, low read count/interaction suggests that these are not true interactions, consistent with the fourSig results. Technical bias estimated by these two approaches is likely conservative – some true biological interactions may be "discounted" in this way, creating false negatives in order to most reliably account for false positive interactions. Despite different underlying assumptions, technical bias estimates derived from the two methods agree with each other to within <1% for both promoter and CTCF site marker SNPs.

In each case, read counts for interactions originating from each haplotype were compared (1) with an expected proportion of 0.5 (in analyses without normalization for technical bias) or (2) with the proportions observed using the estimates of technical bias described above (in analyses normalized for technical bias), using a binomial test.

CTCF site SNP enrichment analysis

We identified all SNPs that have been associated with risk for FTLD, Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis, and related conditions, at a genome-wide significant level ($P \leq 5x10^{-8}$), using the EMBL-EBI/NHGRI GWAS catalog (Welter et al., 2014). This list contains 200 SNPs associated with 29 traits (**Table S4.5**). We downloaded all available optimal irreproducible discovery rate (IDR) thresholded peak (Landt et al., 2012) BED files from ENCODE CTCF ChIP-seq experiments performed in brain-relevant cell/tissue types (https://www.encodeproject.org/; hippocampal astrocytes, cerebellar astrocytes, BE2_C neuroblastoma, retinoic acid-treated SK-N-SH neuroblastoma, choroid plexus epithelial cells, H54 glioblastoma, and brain microvascular endothelial cells). We used the GREGOR pipeline (Schmidt et al., 2015) to determine if the disease-associated SNPs or their LD proxies (at r^2 values of ≥ 0.7, ≥ 0.8 and ≥ 0.9) were enriched in brain CTCF binding sites compared to control SNPs that were matched for minor allele frequency, number of LD proxies, and distance to the nearest gene (Schmidt et al., 2015). 14 of the neurodegenerative risk SNPs were not identifiable by GREGOR, resulting in a final list of 186 disease-associated SNPs and proxies.

CHAPTER 5: DISCUSSION

5.1 Genetic variation at *TMEM106B* affects risk for neurodegeneration

Since the publication of the first genome-wide association study (GWAS) in 2005 (Klein et al., 2005), GWAS have been widely used to identify common genetic risk factors for complex traits and diseases (Edwards et al., 2013). Complex diseases, by definition, do not have single causes (Ghosh and Collins, 1996), which has made it challenging to 1) understand the mechanisms underlying these diseases, and 2) develop therapeutic interventions. Indeed, while thousands of SNP-trait associations have been identified by GWAS (Welter et al., 2014), only a small proportion have been investigated mechanistically (see Chapter 1.2). Such a knowledge gap between 1) known disease-associated loci and 2) the mechanisms by which these loci affect disease risk also exists for neurodegenerative diseases, such as Alzheimer's disease (AD), Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS), and frontotemporal lobar degeneration (FTLD). Specifically, while GWAS approaches have identified ~200 variants that are associated with the various neurodegenerative diseases (Welter et al., 2014), few novel biological insights from these loci have been gained. There are currently no disease-modifying treatments for these fatal diseases, and the prevalence of neurodegenerative disease is expected to increase rapidly as the world population ages (Chen and Zheng, 2012; Katsnelson et al., 2016). Thus, there is an urgent need for advances in our understanding of their underlying pathophysiology.

In 2010, common genetic variants at the *TMEM106B* locus on chromosome 7p21 were found to be significantly associated with risk for FTLD by GWAS (Van Deerlin et al., 2010). In addition, these variants have been associated with various other neurological and neurodegenerative phenotypes, suggesting that genetic variation at the *TMEM106B* locus plays an important role in the pathophysiology of neurodegeneration. As FTLD is an untreatable, fatal disease (Bang et al., 2015; Seelaar et al., 2011), it is important to understand how genetic variation at the *TMEM106B* locus affects disease risk, which is the focus of Chapter 5.2. In this section, I

discuss the discovery and replication of FTLD-associated variants on chromosome 7p21, and the subsequent discovery of the 7p21 locus as a genetic modifier of Mendelian forms of FTLD.

5.1.1 Genetic variation at *TMEM106B* affects risk for frontotemporal lobar degeneration

The first GWAS performed to identify genetic risk factors for FTLD consisted of a neuropathologically homogeneous cohort of FTLD-TDP cases and ancestry-matched controls (Van Deerlin et al., 2010). Since neuropathologically distinct forms of FTLD (FTLD-TDP, FTLD-tau or FTLD-FUS) may result from different underlying disease mechanisms (Mackenzie and Neumann, 2016), the homogeneity of this cohort may have increased power to detect genetic risk factors of relatively small effect (Van Deerlin et al., 2010). Of >500,000 genotyped single nucleotide polymorphisms (SNPs), three SNPs reached a genome-wide significant association with FTLD-TDP. All three SNPs are located in or near the *TMEM016B* gene, which was completely uncharacterized at the time of the discovery (Van Deerlin et al., 2010). The SNPs are in strong linkage disequilibrium (LD) with each other (pairwise r^2=0.7 to 1), and are also in strong LD with >100 other genetic variants, together comprising a 36kb haplotype block in individuals of Western European ancestry (1000 Genomes Project Consortium et al., 2010). The lack of association of these other strongly linked variants with disease is due to the absence of these variants on the genotyping arrays used in the study, which is a common problem in GWAS, since genotyping arrays only genotype ~5% of known human SNPs (Edwards et al., 2013). Thus, the three FTLD-TDP-associated SNPs are simply tagging a much larger FTLD-TDP risk haplotype, and theoretically, any one (or more) of the 100+ variants in strong LD may be the cause of the association (the "causal variant(s)"). The risk alleles of these linked variants have allele frequencies of ~60% in populations of Western Europeans ancestry (1000 Genomes Project Consortium et al., 2015) (**Figure 1.2**), suggesting that the causal variant or variants may be very common. Interestingly, the association of *TMEM106B* SNPs with FTLD-TDP was stronger in *GRN*+ FTLD-TDP cases than in FTLD-TDP cases without *GRN* mutations (Van Deerlin et al., 2010), suggesting a potential genetic modifier effect of *TMEM106B* haplotype in the presence of highly penetrant *GRN* mutations.

Importantly, the association of two of the three SNPs with FTLD-TDP was replicated in an independent cohort in the original GWAS study (one SNP, rs6966915, could not be successfully genotyped in the replication phase) (Van Deerlin et al., 2010), and was also replicated by two independent research groups using largely clinically-defined FTLD cases (Finch et al., 2011; van der Zee et al., 2011). One replication study in particular confirmed that the association was strongest in *GRN*+ FTLD-TDP cases (Finch et al., 2011). A third replication study, Rollinson et al. (2011), did not detect an association between the *TMEM106B* SNPs and FTLD in a cohort of 312 clinical FTLD cases and 248 controls (Rollinson et al., 2011a). However, this case cohort did not include any *GRN* mutation carriers or autopsy-confirmed TDP-43 cases, and thus was likely more heterogeneous (and thus underpowered) than the other studies. It is worth mentioning, however, that this study did observe a trend towards association for one of the three GWAS SNPs, in the same direction reported in the GWAS and other replication studies (Rollinson et al., 2011a). Furthermore, a recent meta-analysis combining the GWAS and all published replication studies confirmed a statistically significant association between rs1990622 and FTLD, with all studies demonstrating the same direction of effect; specifically, the major allele of rs1990622 associates with increased risk for FTLD (Hernandez et al., 2015). Thus, *TMEM106B* is a validated risk locus for FTLD, and appears to be particularly strongly associated with disease risk when analyzing only *GRN*+ FTLD-TDP cases, or FTLD-TDP cases more generally.

A second FTLD GWAS was published in 2014, but rather than using a neuroapthologically-defined cohort, as in the FTLD-TDP GWAS, the authors used multiple cohorts defined by their clinical symptoms (Ferrari et al., 2014). Specifically, the cohorts consisted of anywhere from 115 to 1,634 individuals with FTLD, all with a specific clinical subtype, and 7,444 controls. As with the FTLD-TDP GWAS, all individuals were of Western European ancestry. The study identified significant associations between genetic variants at the *HLA-DRA/HLA-DRB* and *BTNL2* loci with FTLD (no significant sub-type-specific associations were found), but did not observe a significant association with variants at or near *TMEM106B* (Ferrari et al., 2014). However, the use of clinically-defined cases very likely increases neuropathological heterogeneity (see Chapter 1.1), and thus

might preclude the ability to detect a risk factor that affects risk for a specific neuropathological subtype, e.g. *TMEM106B* for FTLD-TDP. Indeed, clinical FTLD subtypes are generally not good predictors of FTLD neuropathological subtypes (Seelaar et al., 2011). In addition, the authors excluded *GRN*+ FTLD-TDP cases, which, as discussed earlier, is a major driver of the association between *TMEM106B* haplotype and FTLD-TDP (Finch et al., 2011; Van Deerlin et al., 2010). Thus, it is not surprising that this GWAS did not detect an association at the *TMEM106B* locus.

Interestingly, the most significant (sentinel) GWAS SNP, rs1990622, has also been associated with various other neurological and neurodegenerative phenotypes. Shortly after the publication of the FTLD-TDP GWAS, the risk allele of rs1990622 was found to be associated with increased risk of cognitive impairment in ALS (Vass et al., 2011), a disease that shares TDP-43 pathology with FTLD-TDP (Neumann et al., 2006). Two other studies demonstrated that the risk allele was associated with reduced plasma levels of PGRN (Cruchaga et al., 2011; Finch et al., 2011), the protein encoded by the *GRN* gene, thus providing further evidence of an epistatic interaction between *TMEM106B* genotype and *GRN*. Since *GRN* mutations cause FTLD through loss-of-function mechanisms (Kleinberger et al., 2013; Petkau and Leavitt, 2014), the risk haplotype at *TMEM106B* may increase disease risk in part by altering PGRN levels. Subsequent studies demonstrated an association between the rs1990622 risk allele and more severe TDP-43 pathology in the brain, both in neurologically normal individuals (Yu et al., 2015) and in individuals with AD (Rutherford et al., 2012), as well as increased risk of hippocampal sclerosis of aging, a common neurodegenerative condition (Nelson et al., 2015). Finally, the rs1990622 risk allele associates with decreased brain functional network connectivity in asymptomatic *GRN* mutation carriers (Premi et al., 2014) and reduced temporal lobe gray matter volume, anterior commissure cross-sectional area, and posterior corpus callosum size (Adams et al., 2014). Thus, the FTLD-TDP risk haplotype at *TMEM106B* affects not only molecular and cellular phenotypes related to neurodegeneration, but macroscopic brain phenotypes as well.

In summary, common genetic variation at the *TMEM106B* locus is significantly associated with risk for developing FTLD-TDP, as well as several other related neurological and

neurodegenerative phenotypes. Furthermore, the risk allele of rs1990622, the GWAS sentinel SNP, is highly overrepresented in *GRN*+ FTLD-TDP cases, suggesting that *TMEM106B* genotype may 1) modify disease in the presence of pathogenic *GRN* mutations, and 2) affect disease risk at least partially through effects on *GRN*. However, these associations do not indicate which genetic variant or variants are causing the associations, nor how one or more functional variants might be acting in order to influence disease risk. Indeed, it is not clear from the association whether *TMEM106B* is actually the gene responsible for conferring disease risk, or whether variation at the *TMEM106B* locus might be affecting disease risk through effects on other genes, e.g. through long-range regulatory interactions. These questions, which are of significant scientific and clinical importance, are discussed in Chapter 5.2.

5.1.2 Genetic variation at *TMEM106B* affects clinical prognosis of Mendelian frontotemporal lobar degeneration

The enrichment of the rs1990622 risk allele in *GRN*+ FTLD-TDP cases, as described above, was initially surprising, given that *GRN* mutations are highly penetrant (Cruts et al., 2006; Gass et al., 2006). Why would a disease risk allele that exerts only a small effect on disease risk (the odds ratio for the risk allele of rs1990622 is 1.64), be overrepresented (or underrepresented) in individuals with a Mendelian disease-causing mutation? One potential explanation is that *TMEM106B* genotype plays a disease-modifying role in the presence of pathogenic *GRN* mutations. Such a mechanism is plausible, given the extremely wide range of ages at which *GRN* mutation carriers develop FTLD symptoms (Benussi et al., 2015). Consistent with this hypothesis, two studies published shortly after the FTLD-TDP GWAS reported that the rs1990622 risk allele is associated with a ~12-13 year earlier age at disease onset in *GRN*+ FTLD-TDP cases (Cruchaga et al., 2011; Finch et al., 2011). An epistatic interaction between *TMEM106B* and *GRN* is further supported by the findings summarized in Chapter 5.1.1, as well as the cellular co-localization of TMEM106B and PGRN within LAMP-1+ organelles (Chen-Plotkin et al., 2012), which mark late endosomes and lysosomes (Lamb et al., 2013). Thus, *TMEM106B* genotype not only affects risk of developing FTLD-TDP, but also affects the clinical presentation of a Mendelian form of FTLD-

TDP. Finch et al. (2011) reported that while 19.1% of control individuals are homozygous for the rs1990622 protective allele, only 2.6% of *GRN*+ FTLD-TDP individuals have this specific genotype (*P*=0.009) (Finch et al., 2011). While it is unclear whether the protective haplotype simply delays age at onset or reduces the penetrance of the *GRN* mutations, these results suggest an interaction between *TMEM106B* and *GRN* in FTLD-TDP pathogenesis.

As discussed in Chapter 1, the three major genetic causes of FTLD are mutations in the *MAPT*, *GRN*, and *C9orf72* genes, with *C9orf72* being the most common cause of FTLD, ALS, and FTLD/ALS (Tan et al., 2017). Importantly, disease-causing hexanucleotide repeat expansion (HRE) mutations result in TDP-43 pathology (Al-Sarraj et al., 2011), and, similar to *GRN* mutations, are likely not fully penetrant, given that they are present in some sporadic FTLD cases (Tan et al., 2017) and have been found in neurologically normal individuals of advanced age (Beck et al., 2013; Harms et al., 2013; Majounie et al., 2012). Thus, it is possible that *TMEM106B* haplotype affects the clinical manifestation of *C9orf72*-associated FTLD (*C9orf72*+ FTLD-TDP) in addition to *GRN*+ FTLD-TDP.

To test this hypothesis, I investigated whether there is an association between rs1990622 genotype and age at onset or age at death in *C9orf72*+ FTLD-TDP cases (DeJesus-Hernandez et al., 2011; Renton et al., 2011). In both discovery and replication cohorts, I observed a statistically significant association between rs1990622 and both age at death and age at disease onset. Surprisingly, each copy of the rs1990622 risk allele *delayed* age at death and age at disease onset, which is opposite the direction seen in *GRN*+ FTLD-TDP (Gallagher et al., 2014). In other words, in the presence of pathogenic *C9orf72* HREs, the genotype that increases risk for FTLD-TDP more generally is actually *protective* against disease manifestation.

These results were specific to *C9orf72*+ FTLD-TDP, as rs1990622 genotype had no effect on age at death or age at onset in mutation-negative FTLD-TDP. At the same time, I also observed a significant enrichment of the rs1990622 risk allele in *C9orf72* HRE carriers, and confirmed the previously published enrichment of the risk allele in *GRN*+ and mutation-negative FTLD-TDP patients (Gallagher et al., 2014). Importantly, an independent study from another group confirmed

the enrichment of the rs1990622 risk allele in *C9orf72*+ FTLD-TDP (van Blitterswijk et al., 2014). This group did not observe an effect of rs1990622 genotype on age at disease onset, but the authors performed their analyses by combining *C9orf72*+ FTLD and ALS cases, thus potentially precluding the ability to detect a disease-specific effect (van Blitterswijk et al., 2014).

As mentioned above, the protective effects of the rs1990622 risk allele in *C9orf72*+ FTLD-TDP seem counterintuitive, especially given that the risk allele is enriched in *C9orf72* HRE carriers. How could a particular genetic variant (representing a particular haplotype) *increase* risk of developing disease in large cohorts consisting of both Mendelian and non-Mendelian cases (Finch et al., 2011; Van Deerlin et al., 2010; van der Zee et al., 2011), while at the same time *protect* against the manifestation of disease in individuals with a specific disease-causing mutation? In the case of *GRN* mutation carriers, the enrichment of the rs1990622 risk allele is consistent with the earlier age at disease onset seen in risk allele carriers – an individual with a *GRN* mutation is more likely to develop FTLD-TDP at any given age if he/she has the *TMEM106B* risk haplotype, thus creating ascertainment bias in the clinic. Indeed, the scarcity of protective allele homozygotes among *GRN*+ FTLD-TDP cases (Finch et al., 2011) suggests that some individuals with the protective haplotype might never present with clinical FTLD symptoms. However, this type of effect cannot explain the *C9orf72* result, in which the rs1990622 risk allele is 1) enriched in *C9orf72*+ FTLD-TDP cases, but 2) appears to delay both age at disease onset and age at death.

Several possible explanations may be compatible with these results. First, we may be observing an artifact of limited sample size. We note, however, that our study, with 30 clinical sites around the world, is very large (and likely the largest possible in this disease), and the differences in local clinical practice would tend to bias towards a negative result. Alternatively, the enrichment of the rs1990622 risk allele in *C9orf72* HRE carriers, combined with the protective effects of the risk allele with regards to disease manifestation, may indicate that the *TMEM106B* FTLD-TDP risk haplotype confers a fitness advantage in individuals with the *C9orf72* HRE.

While epistatic interactions have been reported for other Mendelian diseases, specifically cystic fibrosis (Cutting, 2010) and sickle cell disease (Sankaran et al., 2010), the interaction

between *TMEM106B* and *C9orf72* appears to be a rarer case of "sign epistasis", in which the phenotypic outcome of the interaction depends on genetic background (Weinreich et al., 2005). Specifically, the *TMEM106B* risk haplotype appears to be deleterious in the absence of the *C9orf72* HRE (thus, its association with increased risk of FLTD-TDP), whereas in the presence of the *C9orf72* HRE, it is beneficial. Interestingly, recent cell biological experiments performed in my lab support a sign epistatic relationship for *TMEM106B* and *C9orf72*. As will be discussed in more detail in Chapter 5.3, overexpression of *TMEM106B* in various cell types leads to enlarged LAMP-1+ late degradative organelles. These organelles fail to acidify properly, resulting in impaired protein degradation and cell death (Brady et al., 2013; Busch et al., 2016; Chen-Plotkin et al., 2012; Stagi et al., 2014; Suzuki and Matsuoka, 2016). However, siRNA-mediated knockdown of *C9orf72* rescues these phenotypes (Busch et al., 2016), and it is worth noting that human *C9orf72* HRE carriers have reduced C9orf72 protein levels, presumably because the expanded allele cannot be efficiently translated into functional protein (Ciura et al., 2013; DeJesus-Hernandez et al., 2011; Donnelly et al., 2013; Gijselinck et al., 2012; Tran et al., 2015). These results suggest that both *TMEM106B* and *C9orf72* play roles in lysosomal function, and that in the presence of *C9orf72* HREs, which reduce C9orf72 protein levels, the *TMEM106B* risk haplotype (and resulting increased TMEM106B levels, see Chapter 5.2) may help restore lysosomal homeostasis. In summary, evidence from human genetics, as well as cellular biological experiments, support a disease-modifying role of *TMEM106B* in *C9orf72*+ FTLD-TDP.

Importantly, the association between rs1990622 genotype and age at onset and/or age at death in both *C9orf72*+ and *GRN*+ FTLD-TDP also strongly suggests that the causal variant influencing disease risk and manifestation at the *TMEM106B* locus is common, since these genetic modifier effects were observed in cohorts of only ~50-100 individuals (Cruchaga et al., 2011; Finch et al., 2011; Gallagher et al., 2014). As mentioned above, the *TMEM106B* risk SNP alleles are very common in the population; thus, a causal variant in strong LD with the GWAS risk SNPs could potentially explain these genetic modifier effects, whereas a rare causal variant could not. This is an important point, since an alternative interpretation of GWAS is that common variants may

associate with disease risk as a result of being in low or partial LD with a rare causal variant of strong effect (Gibson, 2012). Thus, it can be difficult to determine *a priori* whether a GWAS causal variant is in strong LD with the disease-associated SNPs (which is typically assumed, but not proven), or driven by a "synthetic association" due to partial LD with a rare causal variant. However, since the associations of *TMEM106B* SNPs with FTLD-TDP 1) risk and 2) age at onset and death are likely driven by the same functional variant(s), these genetic modifier effects effectively rule out the possibility of a rare causal variant at this locus.

5.2 Increased *TMEM106B* levels are implicated in frontotemporal lobar degeneration

The discovery of SNPs at the *TMEM106B* locus that are associated with risk for FTLD-TDP, clinical prognosis in Mendelian forms of FTLD-TDP, several other neurodegenerative phenotypes, and macroscopic brain phenotypes, raises two critical questions: 1) which genetic variant or variants are responsible for influencing these phenotypes, and 2) what are the mechanisms involved? To summarize Chapter 1.2, the challenges in determining the mechanisms underlying a disease-associated risk locus result mainly from two phenomena: first, genetic variants associated with disease by GWAS are typically in strong LD with tens, hundreds, or even thousands of variants that were not genotyped in the GWAS study – thus, the association does not tell you which variant is *causing* the association (the "causal variant") (Edwards et al., 2013). Second, the association alone does not tell you *how* the causal variant is acting to influence disease risk. For example, in the case of the *TMEM106B* FTLD-TDP risk locus, the fact that the disease-associated SNPs are located at or near *TMEM106B* does not necessarily mean that the causal variant is affecting *TMEM106B*, rather than any other gene in the genome. Furthermore, the association does not reveal which molecular mechanisms might be affected by the causal variant, e.g. altering the amino acid sequence of *TMEM106B,* altering the genetic regulation of *TMEM106B* and/or other genes, etc.

For these reasons, the interpretation of disease-associated loci is challenging, and has lagged far behind the pace of the discovery of SNP-trait associations by GWAS, which now number in the thousands (Welter et al., 2014) (see Chapter 1.2). However, in order for a GWAS-identified risk locus to improve our understanding of the pathophysiology of a particular disease, the molecular and cellular mechanisms involved must be determined. Unfortunately, while ~200 SNPs have been associated with risk for AD, PD, ALS, FTLD, and related phenotypes (Welter et al., 2014), only the PD *SNCA* risk locus has been investigated mechanistically with regards to the underlying genetic basis of disease risk (Soldner et al., 2016), and the *SNCA* gene was already known to play a key role in PD (Polymeropoulos et al., 1997; Spillantini et al., 1997). Thus, further

studies are required to determine the mechanisms by which validated risk loci act to confer risk for neurodegenerative disease, as well as for complex traits in general.

In the case of the 7p21 FTLD-TDP risk locus, two important lines of evidence suggest that *TMEM106B* may be the gene responsible for conferring disease risk; specifically, the risk alleles of the GWAS SNPs, as well as other variants in strong LD with them (i.e. the risk haplotype), are significantly associated with increased *TMEM106B* mRNA levels in human cell lines and brain (GTEx Consortium, 2015; Lappalainen et al., 2013; Liang et al., 2013; Stranger et al., 2012; Yu et al., 2015), and this expression quantitative trait locus (eQTL) effect does not exist for any other transcripts genome-wide (GTEx Consortium, 2015; Liang et al., 2013). Thus, the FTLD-TDP causal variant(s) may increase disease risk by increasing *TMEM106B* expression. These findings are consistent with the widely accepted notion that many, if not most, loci associated with disease by GWAS act to influence disease risk through effects on gene expression levels (Maurano et al., 2012; Schaub et al., 2012). Indeed, of the several dozen published studies that have experimentally identified putative GWAS causal variants, nearly all of them are non-coding and affect gene expression levels (see Chapter 1.2). In addition, *TMEM106B* expression levels are increased in FTLD-TDP brains, compared to controls (Chen-Plotkin et al., 2012; Van Deerlin et al., 2010). Thus, two independent lines of evidence suggest that increased *TMEM106B* levels may play a role in FTLD-TDP pathogenesis.

In the next section, I will discuss the discovery of an upstream genetic regulator that appears to be at least partially responsible for the elevated levels of TMEM106B seen in FTLD-TDP brains. In Chapter 5.2.2, I will discuss the results of an unbiased, comprehensive functional characterization of all candidate causal variants at the 7p21 FTLD-TDP risk locus, resulting in the identification of a single SNP that appears to affect transcription factor binding, long-range chromatin interactions, and transcriptional regulation at *TMEM106B*.

5.2.1 TMEM106B expression is elevated in the brains of frontotemporal lobar degeneration patients

After the discovery and replication of genetic variation at *TMEM106B* as a risk factor for FTLD-TDP (Finch et al., 2011; Rollinson et al., 2011a; Van Deerlin et al., 2010; van der Zee et al., 2011), initial follow-up studies investigated the potential role of *TMEM106B* in FTLD-TDP pathogenesis. In the original GWAS, it was observed that *TMEM106B* mRNA levels were elevated in FTLD-TDP brains, particularly in *GRN*+ FTLD-TDP cases, compared to neurologically healthy controls (Van Deerlin et al., 2010). This effect, reported initially in prefrontal cortex samples, was independent of *TMEM106B* genotype, and was later confirmed in temporal and occipital cortex samples (Chen-Plotkin et al., 2012). Elevated TMEM106B protein levels were also confirmed in a small sample of *GRN*+ FTLD-TDP prefrontal cortex samples, compared to controls (Chen-Plotkin et al., 2012).

These observations led to the hypothesis that increased TMEM106B levels in brain may either contribute to disease, result from disease, or both. While I discuss in detail the potential pathogenic role of elevated TMEM106B expression levels in Chapter 5.3, here I discuss the identification of an important *TMEM106B*-regulating microRNA that may at least partially explain the increased TMEM106B levels seen in FTLD-TDP patient brains (**Figure 5.1**).

5.2.1.1 Dysregulation of the microRNA-132/212 cluster contributes to elevated TMEM106B levels in frontotemporal lobar degeneration

The observation that *TMEM106B* levels are elevated in FTLD-TDP, irrespective of *TMEM106B* genotype (Van Deerlin et al., 2010), suggests that one or more upstream pathways regulating the expression of *TMEM106B* is altered in disease. Taking into account the deleterious effects of elevated TMEM106B levels in cell culture-based experiments (Brady et al., 2013; Busch et al., 2016; Chen-Plotkin et al., 2012; Suzuki and Matsuoka, 2016), identifying upstream regulators of *TMEM106B* may reveal potential therapeutic targets.

At the time the GWAS was published, it was becoming increasingly clear that microRNAs (miRs) are critical regulators of gene expression levels, and play important roles in many cellular

processes (Ameres and Zamore, 2013; Jonas and Izaurralde, 2015). With regards to the nervous system, miRs have been implicated in neurite outgrowth, neuronal differentiation, synaptic plasticity, and many other pathways. In addition, miR dysfunction has been implicated in several neurodegenerative diseases, including AD, PD, ALS, and Huntington's disease (Goodall et al., 2013)

My lab performed a microarray-based screen to quantify all known human miRs in FTLD-TDP and neurologically normal brains, and identified miR-132 as the most dysregulated miR. Specifically, miR-132 is expressed at lower levels in FTLD-TDP brains (Chen-Plotkin et al., 2012), and miR-132 expression has been shown to be required for learning, memory, and neuronal dendritic branching (Aten et al., 2016). Furthermore, miR-132 has also been shown to be downregulated in AD, Huntington's disease, and schizophrenia (Aten et al., 2016); thus, miR-132 may be protective against neurological disease more generally. In addition, two other miRs that are processed from the same primary transcript as miR-132, miR-132* and miR-212 (Vo et al., 2005), were also significantly downregulated in FTLD-TDP brains (Chen-Plotkin et al., 2012), suggesting that the locus from which they arise is less transcriptionally active in disease.

Surprisingly, out of 283 predicted target genes of miR-132 and miR-212 (both miRs have the same "seed", and thus many overlapping targets), *TMEM106B* was the top predicted target by TargetScan (Lewis et al., 2003), an online microRNA target prediction tool. I confirmed this prediction using several complementary cell-based assays, specifically demonstrating that miR-132 and miR-212 repress *TMEM106B* mRNA and protein levels through two binding sites in its 3'UTR (Chen-Plotkin et al., 2012). Since miR-132 is expressed at ~100-fold higher levels than miR-212 (Chen-Plotkin et al., 2012; Magill et al., 2010), it is likely more functionally important. Therefore, reduced levels of miR-132 in FTLD-TDP may contribute to the elevated levels of TMEM106B seen in disease brains (**Figure 5.1**).

While the identification of these miRs as regulators of *TMEM106B* is notable, several questions remain: first, are there other, more important regulators of *TMEM106B* that contribute to its upregulation in disease? Indeed, the most recent version of TargetScan identifies 287 predicted

miR binding sites throughout *TMEM106B,* most of which have different predicted miR regulators (Agarwal et al., 2015). While it is unlikely that all of these predicted target sites are truly functional, this does indicate that there are likely multiple miRs that directly regulate *TMEM106B* levels. Notably, the therapeutic potential of targeting miR-132 is questionable, since both overexpression and knockdown approaches have impaired neurological function *in vivo* (Hansen et al., 2010; Magill et al., 2010; Scott et al., 2012; Wayman et al., 2008). In addition, there are likely many transcriptional pathways that may be affected in FTLD-TDP that either directly or indirectly affect *TMEM106B* levels. In support of this, the ENCODE Project has identified >70 transcription factors that bind (and presumably regulate) the *TMEM106B* promoter (Gerstein et al., 2012), some of which may themselves be alternatively regulated in FTLD-TDP. Second, does the downregulation of miR-132/212 and upregulation of *TMEM106B* contribute to disease, result from disease, or both? If this is truly a key pathway in FTLD pathogenesis, animal models may be required in order to determine the specific role of these miRs, as well as TMEM106B levels, in disease pathogenesis. For example, miR-132 is known to be induced by synaptic activity, in which BDNF signaling activates CREB target genes (including miR-132/212) in order to facilitate neuronal plasticity (Vo et al., 2005). Do defects in this signaling pathway contribute to the development of FTLD by altering miR-132-mediated regulation of *TMEM106B*, and, potentially, other genes? Or, conversely, does the presence of FTLD adversely affect these pathways? It is also possible that increased TMEM106B levels may act upstream of miR-132/212 dysregulation. For example, given the deleterious effects of increased TMEM106B levels in cells (see Chapter 5.3), elevated TMEM106B levels may first accelerate disease pathways leading to neurodegeneration, which then results in impaired miR-132/212 regulatory pathways, possibly due to impaired BDNF signaling and/or synaptic plasticity. In this way, *TMEM106B* upregulation may be reinforced through a positive feedback loop involving impaired repression by miR-132/212.

In Chapter 5.3, I discuss the evidence supporting a deleterious effect of increased TMEM106B levels on cells. Before that, however, I will discuss evidence that more than one route may lead to the same intermediate step of increased *TMEM106B* expression (**Figure 5.1**).

Specifically, I will discuss the primary data presented in Chapter 4, suggesting that the common variant responsible for the association of the 7p21 locus with FTLD-TDP risk affects *TMEM106B* expression levels through changes in chromatin architecture.

5.2.2 The *TMEM106B* risk haplotype is associated with increased *TMEM106B* levels

Chapter 1.2 discusses in detail the challenges regarding interpreting the biological mechanisms underlying the association of a genetic locus with disease risk by GWAS. These challenges were re-stated above, in Chapter 5.2, regarding 1) the concept of a haplotype block, in which many co-inherited variants are often similarly associated with disease risk or other phenotypes (Edwards et al., 2013), and 2) the difficulty in determining the molecular mechanisms by which the "causal variant" influences disease risk, due in large part to the location of most disease-associated variants in non-protein-coding regions of the genome (Maurano et al., 2012; Schaub et al., 2012). Consequently, the questions that arise after the identification of a disease-associated locus are: 1) what is the causal variant, and 2) how does the causal variant act to influence disease risk?

Fortunately, a plausible hypothesis regarding question 2 can sometimes be formulated based on publicly available genome-wide studies. As discussed in Chapter 1.2, thousands of genetic variants have been associated with gene expression levels in an unbiased, genome-wide manner, in dozens of cell and tissue types (Albert and Kruglyak, 2015). These expression quantitative trait loci (eQTLs) show significant overlap with GWAS-identified risk variants (Fu et al., 2012; Nicolae et al., 2010; Schaub et al., 2012), suggesting that some disease risk loci may act by altering the expression levels of one or more genes. This result is consistent with the observation that >90% of disease-associated variants reside in non-protein-coding regions of the genome, and are enriched in predicted *cis*-regulatory elements, as determined by epigenomic features such as chromatin accessibility, transcription factor (TF) binding, and specific post-translational histone modifications (Farh et al., 2015; Maurano et al., 2012; Raj et al., 2014; Schaub et al., 2012).

In the FTLD-TDP GWAS, Van Deerlin and colleagues (2010) reported that the risk alleles of the FTLD-TDP-associated SNPs had been associated with increased *TMEM106B* mRNA levels

in an eQTL study performed in Epstein-Barr virus-immortalized B lymphoblastoid cell lines (LCLs) (Dixon et al., 2007; Van Deerlin et al., 2010). This result has been replicated several times (Lappalainen et al., 2013; Liang et al., 2013; Stranger et al., 2012), and has also been reported in various primary immune cell types (Fairfax et al., 2014; Peters et al., 2016; Raj et al., 2014), as well as several human brain regions (GTEx Consortium, 2015; Yu et al., 2015). Furthermore, the GWAS SNPs are not significantly associated with the expression levels of any other genes (GTEx Consortium, 2015; Liang et al., 2013). Thus, correlative data suggests that the *TMEM106B* risk haplotype may increase disease risk by increasing *TMEM106B* levels. The establishment of an eQTL effect also helps one begin to address question 1 above, since the causal variant underlying such an effect is likely to be located in either a transcriptional regulatory element or a region of the *TMEM106B* transcript that affects mRNA splicing or stability.

In the following section, I discuss the results of the functional characterization of all candidate causal variants at the *TMEM106B* FTLD-TDP risk locus, as described in Chapter 4. By combining trans-ethnic fine-mapping, bioinformatics, experimental approaches, and computational analyses, I identify a putative noncoding causal variant that affects binding of the transcription factor CCCTC-binding factor (CTCF) ~7kb downstream of *TMEM106B*, with resulting alterations in long-range chromatin interactions between multiple *cis*-regulatory elements. I also demonstrate that SNPs associated with risk for neurodegenerative diseases in general are enriched in brain CTCF binding sites, suggesting a more general role of CTCF-mediated gene expression programs in neurodegeneration. These results provide an example of the functional dissection of a GWAS locus, which has rarely been done in the field of neurodegeneration. I highlight below the key strengths and weaknesses of our approach, comparing them with other examples from the rapidly-developing field of functional genomics.

5.2.2.1 A frontotemporal lobar degeneration-associated risk variant near *TMEM106B* affects chromatin architecture and *TMEM106B* regulation

The FTLD-TDP risk haplotype spans ~36kb and contains >100 strongly linked variants, including the three SNPs associated with disease risk by GWAS. In order to narrow down the list

of candidate causal variants, I performed trans-ethnic fine-mapping of the *TMEM106B* eQTL signal with a multi-ethnic LCL eQTL study (Stranger et al., 2012). Trans-ethnic fine-mapping takes advantage of the smaller haplotype blocks present in African populations (Campbell and Tishkoff, 2008), and has successfully been used to refine regions of genomic association first identified in individuals of Western European or Asian ancestry (Edwards et al., 2013; Glubb et al., 2015). In addition, I used this multi-ethnic study to perform conditional analyses in order to demonstrate that the association signals at the *TMEM106B* locus for 1) *TMEM106B* expression and 2) disease risk are statistically indistinguishable. Thus, the same causal variant(s) likely underlie both signals, suggesting that the eQTL effect is indeed the molecular phenotype that confers disease risk at this locus.

I then investigated the functions of all three predicted LCL *cis*-regulatory elements (CREs) that contain at least one of the fine-mapped eQTL variants. Two of the CREs display epigenetic features associated with inactive or "poised" enhancers in LCLs (Creyghton et al., 2010; Rada-Iglesias et al., 2011); consistently, they showed little to no enhancer activity in LCL reporter assays, regardless of which SNP alleles were present. Furthermore, these regions do not display any activity-associated epigenetic marks in the various brain tissues analyzed by the Roadmap Epigenome Project (Roadmap Epigenomics Consortium et al., 2015), suggesting that they are not functional in brain. While there are several potential reasons that these results may be false negatives, they are relatively unlikely: first, while some enhancers only become active in response to certain cellular signals (Engel et al., 2016; Shlyueva et al., 2014), the *TMEM106B* eQTL effect has reproducibly been detected in unstimulated LCLs (GTEx Consortium, 2015; Lappalainen et al., 2013; Liang et al., 2013; Stranger et al., 2012). Thus, the mechanisms underlying the eQTL effect should be present in unstimulated LCLs. Second, plasmid-based reporter assays test genomic regions for regulatory activity in an artificial context. Many genomic features, such as association with histones, the relative positions of genomic regions, and long-range chromatin interactions, are absent in the context of a reporter plasmid; thus, such assays are subject to false positive and false negative results (Inoue and Ahituv, 2015). While the epigenetic signatures of these potential

enhancer regions at the *TMEM106B* locus are not consistent with an active state, genomic deletion of these regions would be required to definitively demonstrate that they have either no effect, or no allele-specific effect, on *TMEM106B* levels.

The third CRE contains three SNPs and displays binding of the transcription factor CTCF in virtually all cell types tested, including LCLs, neuronal and glial cells. CTCF has been implicated in various aspects of eukaryotic gene expression over ~25 years, including transcriptional activation, enhancer blocking, and chromatin barrier functions (Ong and Corces, 2014). Over the past ~10 years, however, genome-wide analyses of CTCF binding sites and chromatin architecture in mammalian cells have begun to establish a more central theme regarding the role of CTCF in gene regulation. These studies strongly suggest that the primary role of CTCF in mammals is to organize the three-dimensional structure of the genome by facilitating long-range chromatin looping interactions (Dixon et al., 2012; Kim et al., 2007b; Nora et al., 2012; Rao et al., 2014; Tang et al., 2015; Zuin et al., 2014). These interactions result in the compartmentalization of the genome into several tiers of organization, of which the best characterized is the "topologically-associated domain", or "TAD", which typically ranges in size from ~40kb to ~3Mb (Dixon et al., 2012; Merkenschlager and Nora, 2016; Nora et al., 2012; Rao et al., 2014). Active genes and enhancer/promoter (E/P) interactions tend to be located within TADs, and several studies have demonstrated that particular CTCF binding sites are required for establishing TADs and enhancer-mediated gene regulation (Dowen et al., 2014; Guo et al., 2015; Lupianez et al., 2015; Rao et al., 2014; Tang et al., 2015). Therefore, I hypothesized that one or more SNPs in this CTCF site might regulate *TMEM106B* transcription through effects on long-range interactions between distal CREs.

In support of this hypothesis, one of the three SNPs located within the CTCF binding region, rs1990620 (which is in complete LD with the sentinel GWAS SNP rs1990622), displays allele-specific recruitment of CTCF in nuclear extract, and these results are corroborated by CTCF ChIP-seq experiments performed in various cell lines. The agreement between these two approaches, which employ completely different methodologies, strengthens the conclusion that rs1990620 functionally affects CTCF binding at this region. Importantly, Capture-C experiments demonstrate

that strong long-range interactions take place between the CTCF site, the *TMEM106B* promoter, and a predicted active enhancer, and that these interactions take place within a ~250kb TAD delineated by CTCF boundary sites. Interestingly, this enhancer was confirmed to produce bidirectional transcripts, a distinguishing feature of active enhancers (Li et al., 2016), in primary human blood cells and brain samples by the FANTOM Consortium (Andersson et al., 2014). Moreover, the transcriptional activity of the enhancer correlates significantly with the transcriptional activity of *TMEM106B* across cell and tissue types (Andersson et al., 2014), which, combined with the Capture-C data, suggest that this enhancer regulates *TMEM106B*. Thus, I hypothesized that rs1990620 might affect *TMEM106B* transcription by altering CTCF-mediated long-range interactions within the TAD, with indirect effects on interactions between *TMEM106B* and a putative active enhancer. Consistently, analysis of the Capture-C data using haplotype marker SNPs indeed suggests that the long-range interactions at the *TMEM106B* locus are stronger on the risk haplotype in heterozygous cells. These data are consistent with the current model of TAD function, in which CTCF sites facilitate E/P interactions within TADs, but prevent E/P interactions between different TADs (Merkenschlager and Nora, 2016).

While several studies have demonstrated allele-specific long-range interactions between putative CREs harboring nominated GWAS causal variants and their target promoters (Dunning et al., 2016; Nakaoka et al., 2016; Painter et al., 2016; Stadhouders et al., 2014; Vicente et al., 2015), these findings, to my knowledge, represent the first example of a GWAS causal variant affecting the higher-order chromatin architecture within which enhancer/promoter interactions take place. Indeed, few studies have investigated allele-specific CTCF-mediated interactions, and the genome-wide approaches used to investigate them drastically reduce the sensitivity to detect them (Tang et al., 2015). Thus, the Capture-C approach utilized in Chapter 4 will likely be an important tool for investigating chromatin architecture in more detail at specific loci.

While the results discussed here are consistent with a role of one particular CTCF site in allele-specific long-range interactions and transcriptional regulation (**Figure 5.1**), several key questions remain unresolved. First, which specific interactions are required for allele-specific

expression of *TMEM106B*? Our experiments lacked the sequencing depth required to detect allele-specific differences in specific interactions, such as between the *TMEM106B* promoter and the CTCF site, or between the promoter and the enhancer. However, the traditional chromosome conformation capture (3C) approach, which employs qPCR primers specific for pre-selected interacting regions, might be sufficiently quantitative to detect such effects (Denker and de Laat, 2016). Second, the magnitude of the allele-specific interaction differences are small: from the promoter viewpoint, I observe a 6-13% difference that is reduced to a 4-8% difference after adjusting for the estimated technical bias; from the CTCF site viewpoint, I observe a 5-15% difference after adjustment for bias, depending on which marker SNP is used. These results are highly significant since they are based on thousands of sequencing reads, but it is unclear whether such small changes in long-range interaction frequencies would be sufficient to cause a meaningful difference in target gene expression levels. As mentioned above, very little is known about the nature of allele-specific long-range interactions. It is also possible that at the *TMEM106B* locus, only one, or a few, specific interactions are truly allele-specific, in which case totaling all the interactions across large genomic intervals, as was done in the analyses presented in Chapter 4, would dilute the estimated magnitude of the effect. Third, the allele-specific differences in long-range interactions could be unrelated to the allele-specific expression of *TMEM106B*, and could be due to either technical or biological noise. Fourth, no experiments (with the exception of EMSAs) were performed in brain-relevant tissue or cell lines. Indeed, while the eQTL effect has been reported in several brain regions (GTEx Consortium, 2015; Yu et al., 2015), it is not clear which cell type(s) the eQTL effect is coming from. I discuss potential follow-up experiments to address these questions in detail in Chapter 5.4.

In summary, I performed a comprehensive prioritization and functional characterization of all candidate causal variants at the 7p21 FTLD-TDP risk locus. The association of the FTLD-TDP risk SNPs with *TMEM106B* expression and no other genes, combined with the fine-mapping and conditional analyses, suggest that one or more causal variants at this locus influence disease risk by altering *TMEM106B* expression. After ruling out effects on mRNA stability and prioritizing

candidate causal variants based on epigenomic annotations in LCLs and brain-relevant cell lines, I identified three potential CREs, only one of which displays evidence of activity in brain. This CRE contains a putative functional SNP that affects recruitment of CTCF, and potentially other TFs, downstream of *TMEM106B*, and this appears to affect long-range regulatory interactions between multiple CREs spanning hundreds of kilobases. While the specific mechanisms linking allele-specific long-range interactions with allele-specific expression of *TMEM106B* have yet to be determined, I also demonstrated that SNPs associated with risk for neurodegenerative diseases are enriched in brain CTCF binding sites, as determined by ENCODE ChIP-seq experiments. Thus, CTCF-mediated gene expression programs may play a more general role in risk for neurodegeneration.

5.3 Increased TMEM106B levels lead to lysosomal abnormalities and cell death

In Chapter 5.2, I summarized two independent mechanisms by which increased TMEM106B levels may be involved in FTLD pathogenesis. First, TMEM106B levels are increased in FTLD-TDP brains, particularly in *GRN*+ FTLD-TDP cases, and this at least partially results from the downregulation of microRNA-132 and microRNA-212, which repress TMEM106B levels directly (Chen-Plotkin et al., 2012; Van Deerlin et al., 2010). Second, the FTLD-TDP risk haplotype is associated with increased *TMEM106B* mRNA levels in human cell lines and brain (GTEx Consortium, 2015; Lappalainen et al., 2013; Liang et al., 2013; Stranger et al., 2012; Yu et al., 2015), and this appears to result from allele-specific recruitment of the transcription factor CTCF to a *cis*-regulatory region downstream of *TMEM106B*. The CTCF binding site, as well as the *TMEM106B* promoter, are involved in long-range interactions between distal *cis*-regulatory elements, and these interactions are stronger on the risk haplotype. Given that several recent studies have implicated such long-range interactions in transcriptional activation, (Merkenschlager and Nora, 2016; Rao et al., 2014; Tang et al., 2015), this may be the mechanism by which *TMEM106B* levels are increased in risk haplotype carriers. While these results suggest a putative mechanism by which disease-associated genetic variation at this locus affects disease risk, they do not demonstrate whether increases in *TMEM106B* levels are deleterious.

In the following sections, I will discuss published and un-published work that suggest that the TMEM106B protein is involved in endolysosomal pathways, and that increased TMEM106B levels result in endolysosomal abnormalities, impaired protein degradation, and cell death (**Figure 5.1**). I will also discuss the role of C9orf72 in modifying these phenotypes, as this may underlie the ability of *TMEM106B* genotype (and thus, genotype-dependent *TMEM106B* expression levels) to modify disease course in *C9orf72*+ FTLD-TDP cases (Gallagher et al., 2014).

5.3.1 Increased TMEM106B levels cause endolysosomal phenotypes and cell death

When the *TMEM106B* locus was first associated with FTLD-TDP by GWAS (Van Deerlin et al., 2010), nothing was known about the *TMEM106B* gene, or the encoded protein. While the GWAS association does not prove that *TMEM106B* is the gene responsible for conferring risk for

disease (Edwards et al., 2013), no other genes are located within ~380kb upstream and ~85kb downstream of *TMEM106B* (Hinrichs et al., 2016). Thus, initial studies focused on characterizing the TMEM106B protein, under the assumption that *TMEM106B* was indeed the gene responsible for the GWAS association. As discussed in Chapter 5.2, the reported associations between the FTLD-TDP-associated SNPs and the expression levels of *TMEM106B*, but no other genes, support this assumption.

Multiple studies have characterized the molecular and cellular roles of TMEM106B, and most have reached similar conclusions. First, TMEM106B is a heat-sensitive, glycosylated, 31kDa protein that exists in both monomeric and dimeric forms, and localizes mainly to late LAMP-1+ organelles, which consist of late endosomes and lysosomes (Brady et al., 2013; Busch et al., 2016; Chen-Plotkin et al., 2012; Lang et al., 2012; Nicholson et al., 2013; Schwenk et al., 2014; Stagi et al., 2014). Thus, TMEM106B may be involved in protein degradation pathways, such as endocytosis and autophagy (Lamb et al., 2013). More specifically, TMEM106B is an integral endo/lysosomal membrane protein oriented in a Type II manner, with the N-terminus facing the cytoplasm (Lang et al., 2012; Stagi et al., 2014). TMEM106B expression has been detected in various human cancer cell lines, mouse and rat primary neurons, and non-human primate fibroblasts, and immunohistochemistry has confirmed TMEM106B expression in both neuronal and non-neuronal cell types in human brain tissue sections (Brady et al., 2013; Busch et al., 2013; Busch et al., 2016; Chen-Plotkin et al., 2012; Lang et al., 2012; Nicholson et al., 2013; Schwenk et al., 2014; Stagi et al., 2014).

Second, TMEM106B co-localizes extensively with PGRN (Brady et al., 2013; Chen-Plotkin et al., 2012; Nicholson et al., 2013), suggesting a potential interaction between these two FTLD-associated proteins (Cruts et al., 2006; Gass et al., 2006). Several studies have reported that overexpression of *TMEM106B* results in the intracellular (and sometimes extracellular) accumulation of PGRN (Brady et al., 2013; Chen-Plotkin et al., 2012; Nicholson et al., 2013). These results are interesting, given the reported associations between *TMEM106B* genotype and plasma PGRN levels, as well as the ability of *TMEM106B* genotype to modify age at disease onset in *GRN*+

FTLD-TDP cases (Cruchaga et al., 2011; Finch et al., 2011). Thus, *TMEM106B* genotype may affect disease manifestation in *GRN*+ FTLD-TDP, as well as risk for FTLD-TDP more generally, through effects on PGRN levels.

Third, manipulation of TMEM106B levels alters lysosomal morphology and function. Many studies have reported that overexpression of *TMEM106B* results in a significant enlargement of LAMP-1+ late endosomes and lysosomes, referred to from here on as lysosomes for simplicity (Brady et al., 2013; Busch et al., 2016; Chen-Plotkin et al., 2012; Stagi et al., 2014), whereas TMEM106B knockdown does not appear to affect lysosomal size (Stagi et al., 2014). Interestingly, the enlarged lysosomes seen in *TMEM106B*-overexpressing cells fail to acidify properly (Brady et al., 2013; Busch et al., 2016; Chen-Plotkin et al., 2012), and we and others have reported impaired degradation of the epidermal growth factor receptor after treatment of cells with its ligand, epidermal growth factor (Brady et al., 2013; Busch et al., 2016). Thus, increased TMEM106B levels may adversely affect protein degradation through effects on the endolysosomal pathway.

Finally, *TMEM106B* overexpression has also been reported to induce cell death through caspase-dependent pathways, and possibly lysosomal pathways as well (Busch et al., 2016; Suzuki and Matsuoka, 2016). TMEM106B-induced cytotoxicity and impaired lysosomal function is dependent upon localization of TMEM106B to lysosomes, as overexpression of a mutant form of TMEM106B that fails to localize to lysosomes does not cause these phenotypes, despite high expression levels (Busch et al., 2016). Furthermore, I demonstrate in Chapter 4 that even modest (~2-fold) increases in TMEM106B levels are sufficient to produce these phenotypes, and I have found that even lower levels (~1.2-fold) of overexpression are sufficient to cause the enlarged lysosome phenotype (unpublished data). These results suggest that the modest expression level changes associated with the *TMEM106B* eQTL effect (estimated magnitudes range from ~15-60%, depending on cell type (GTEx Consortium, 2015; Liang et al., 2013; Peters et al., 2016; Stranger et al., 2012)), which likely underlies the effect of this locus on disease risk, are sufficient to produce deleterious phenotypes in human cells (**Figure 5.1**).

To further investigate the effects of TMEM106B levels on lysosomes in neurons, Schwenk et al. (2014) knocked down TMEM106B in primary rat hippocampal neurons, and observed enhanced retrograde transport of lysosomes and impaired dendritic branching (Schwenk et al., 2014). The effect on dendritic branching was phenocopied by overexpressing a dominant-negative form of Rab7a, which blocks cargo transport from early to late endosomes and lysosomes, and inhibits lysosomal biogenesis (Mukhopadhyay et al., 1997; Press et al., 1998). Thus, TMEM106B may be required for proper dendritic branching through effects on lysosomal function. In order to identify TMEM106B-interacting partners, the authors immunoprecipitated TMEM106B from P15 rat brain, and performed mass spectrometry to identify other proteins that were pulled down along with it. With this approach, they identified microtubule-associated protein 6 (MAP6) as the only TMEM106B-interacting protein found in three independent experiments, subsequently confirming its interaction with TMEM106B with co-immunoprecipitation experiments (Schwenk et al., 2014). Interestingly, overexpression of MAP6 phenocopied the effects of TMEM106B, whereas MAP6 knockdown rescued these effects in TMEM106B knockdown cells (Schwenk et al., 2014). Thus, TMEM106B and MAP6 may play opposing roles in lysosomal transport and dendritic branching.

Stagi et al. (2014) independently investigated potential TMEM106B-interacting partners by performing a yeast two-hybrid screen (Stagi et al., 2014). They performed two separate screens, with one using the cytosolic N-terminal fragment (amino acids 1-96) of TMEM106B as the bait, and the other using the luminal C-terminal fragment (amino acids 118-274) as the bait. While it is possible that splitting the protein into two parts in this manner could result in false positives and false negatives due to effects on secondary and tertiary structure, the authors confirmed interactions identified in the screen with subsequent co-immunoprecipitation experiments. Using this approach, they identified the endocytic adaptor proteins adaptor related protein complex μ 1 subunit (AP2M1) and clathrin heavy chain (CLTC) as TMEM106B N-terminal interacting partners (Stagi et al., 2014). In addition, transmembrane protein 106C (TMEM106C), which shares only 47% amino acid homology with TMEM106B, was also shown to interact with TMEM106B, resulting in heterodimers at the lysosomal membrane; unfortunately, nothing is known about the cellular

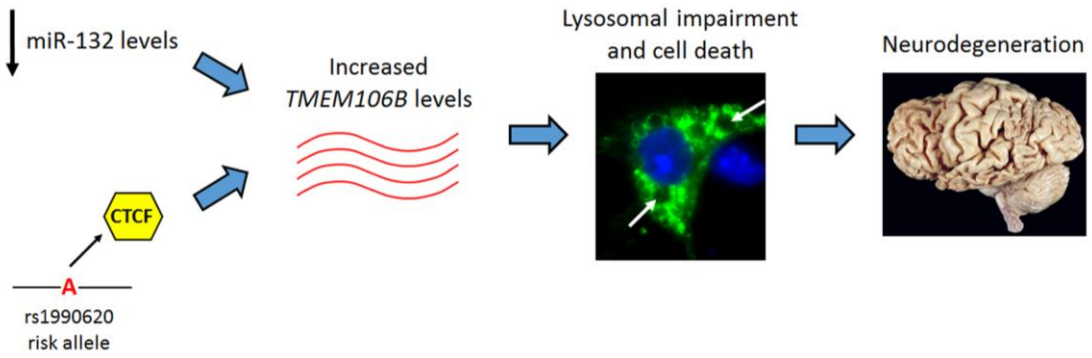


Figure 5.1. Model for how increased *TMEM106B* expression levels increase risk for FTLD.
Increased levels of TMEM106B in the brain result from 1) decreased levels of the negative regulator microRNA-132, and/or 2) the risk allele of rs1990620, which increases CTCF-mediated long-range regulatory interactions. At the cellular level, increased TMEM106B levels impair lysosomal function and cellular health, which increases risk for neurodegeneration and the development of FTLD symptoms.

functions of TMEM106C. While the authors did not determine the functional implications of these interactions, they went on to demonstrate that TMEM106B overexpression and knockdown impaired and enhanced lysosomal transport in both anterograde and retrograde directions, respectively, in mouse cortical neurons (Stagi et al., 2014). They also showed that *TMEM106B* overexpression activates the transcription factor EB (TFEB)-mediated lysosomal stress response, in which TFEB localizes to the nucleus to activate lysosomal stress response genes (Sardiello et al., 2009; Stagi et al., 2014).

In conclusion, multiple studies have confirmed a role for TMEM106B in lysosome function, protein degradation and cellular health. TMEM106B also co-localizes with PGRN and affects PGRN levels, which may underlie the *TMEM106B* genetic modifier effect in *GRN*+ FTLD-TDP cases. While potential *TMEM106B* interacting partners have been proposed, the two relevant studies employed different approaches and obtained non-overlapping results. Taken together, further studies will be required in order to identify the precise molecular function of TMEM106B, and how altered TMEM106B levels affects lysosomal and other cellular phenotypes.

5.3.2 TMEM106B and C9orf72 interact to affect endolysosomal phenotypes and cell death

Given the genetic modifier effects of *TMEM106B* genotype in *GRN*+ FTLD-TDP and *C9orf72*+ FTLD-TDP cases, the molecular and cellular relationships between TMEM106B, PGRN and C9orf72 are an important area of inquiry. As mentioned above, TMEM106B and PGRN co-localize at LAMP-1+ organelles, and TMEM106B overexpression affects PGRN levels. Thus, expression levels of TMEM106B, which are influenced by *TMEM106B* genotype, may influence disease manifestation through effects on PGRN levels.

Based on the discovery that *TMEM106B* genotype also affects the course of disease in *C9orf72*+ FTLD-TDP (Gallagher et al., 2014), my lab has investigated the potential cellular relationship between TMEM106B and C9orf72. This area of investigation is particularly important given the surprising direction of the genetic modifier effect in *C9orf72*+ FTLD-TDP; specifically, the major allele of rs1990622, the sentinel GWAS SNP, is associated with *increased risk* of FTLD-TDP in cohorts consisting of both Mendelian and non-Mendelian cases (Finch et al., 2011; Hernandez et al., 2015; Van Deerlin et al., 2010; van der Zee et al., 2011), but *delays* age at disease onset and age at death in *C9orf72*+ FTLD-TDP cases (Gallagher et al., 2014). Since the pathogenic *C9orf72* hexanucleotide repeat expansion (HRE) mutations are thought to result in both loss of normal *C9orf72* function (as a result of reduced wild-type C9orf72 mRNA and protein levels) and one or more toxic gain-of-functions (Todd and Petrucelli, 2016), the genetic modifier role of *TMEM106B* may be difficult to understand mechanistically.

Recent work from my lab and others, however, has suggested that *TMEM106B* and *C9orf72* may play similar roles in lysosomal function and lysosome-related cellular phenotypes. As mentioned in Chapter 5.1.2, my lab has shown that the deleterious effects of *TMEM106B* overexpression (enlarged, poorly acidified lysosomes, and cytotoxicity) are rescued upon *C9orf72* knockdown, and thus, depend on the presence of C9orf72 (Busch et al., 2016). These experiments suggest that the normal functions of C9orf72 (vs. the aberrant toxic functions conferred by the RNA foci and dipeptide repeat proteins originating from the HRE) may lie in the same lysosomal

pathways as those involving TMEM106B. Indeed, several recent studies have implicated C9orf72 in lysosomal/autophagy pathways (Farg et al., 2014; Sellier et al., 2016; Sullivan et al., 2016; Ugolino et al., 2016; Webster et al., 2016; Yang et al., 2016). Specifically, C9orf72 has been shown to interact with several autophagy-associated proteins, such as RAB1, SMCR8, WDR41, and ULK1 (Ciura et al., 2016; Farg et al., 2014; Jung et al., 2017; Sellier et al., 2016; Sullivan et al., 2016; Webster et al., 2016; Yang et al., 2016); indeed, SMCR8 has been shown to regulate ULK1 kinase activity, with both proteins implicated in autophagosome biogenesis (Jung et al., 2017; Webster et al., 2016; Yang et al., 2016). Moreover, a C9orf72-SMCR8 complex appears to act as a Rab guanine nucleotide exchange factor (GEF) for RAB8 and RAB39 (Ciura et al., 2016; Sellier et al., 2016; Yang et al., 2016), and loss of C9orf72 has been reported to increase starvation-induced autophagic flux (Ugolino et al., 2016).

Taken together, these studies suggest a lysosomal/autophagic function for both TMEM106B and C9orf72. Whether the involvement of normal C9orf72 in lysosomal functions contributes to the development of FTLD-TDP and ALS in *C9orf72* HRE carriers, however, remains unclear.

5.4 Unanswered questions and future directions

In the preceding sections of this dissertation, I have made a case for the importance of CTCF- and miR-132-based mechanisms in the upregulation of *TMEM106B*, with resulting increases in TMEM106B levels increasing risk for FTLD by affecting lysosomal and autophagic pathways. I conclude here by considering alternative interpretations of our data, as well as outlining potential future research directions.

My work suggests that genetic variation at a single causal SNP, rs1990620, contributes to FTLD-TDP risk through effects on the transcriptional regulation of *TMEM106B*. However, it is possible that genetic variation at *TMEM106B* affects disease risk through effects on protein function, rather than gene expression. In such a scenario, the eQTL effect at *TMEM106B* would not be responsible for conferring disease risk, but would rather be an independent phenomenon. While the conditional analyses performed in Chapter 4 are consistent with the eQTL effect being driven by the same causal variant that affects disease risk, they do not rule out the possibility that the two associations might be the result of two different causal variants that are in strong LD. For example, several groups have investigated the potential function of the nonsynonymous SNP rs3173615, which causes a serine (protective allele) to threonine (risk allele) substitution in exon 6 of TMEM106B, and is in complete linkage disequilibrium (LD) (r^2=1) with rs1990622. While this SNP does not appear to affect lysosomal phenotypes or PGRN levels (Brady et al., 2013; Nicholson et al., 2013; Stagi et al., 2014), it may affect other molecular and cellular functions that have not yet been thoroughly characterized. For example, one recent study reported that the risk allele of rs3173615 impairs autophagic flux and protein degradation rates compared with the protective allele (Jun et al., 2015), but these results need to be confirmed by independent groups. Another group has reported that the risk allele of rs3173615 increases TMEM106B monomer stability, and thus may increase steady-state protein levels (Nicholson et al., 2013), but other groups have failed to replicate this effect (Jun et al., 2015). Furthermore, this study focused only on monomers and ignored TMEM106B dimers, which have been reported to exist by multiple groups (Brady et al., 2013; Chen-Plotkin et al., 2012). Indeed, using our well-characterized antibody that recognizes

both TMEM106B monomer and dimer species, I repeated the Nicholson et al. (2013) experiments and found that while rs3173615 may affect TMEM106B dimerization, and thus the dimer to monomer ratio, total steady-state protein levels were unaffected (unpublished data). Future studies are needed to investigate the precise molecular and cellular functions of TMEM106B, and whether this amino acid-altering polymorphism affects any of those functions.

Second, even if *TMEM106B* affects disease risk through the eQTL effect described in Chapters 1 and 4, it is unclear whether one or more SNPs within the CTCF binding site truly alters *TMEM106B* levels, or how this might occur. To further demonstrate that this CTCF site affects *TMEM106B* expression in an allele-specific manner, one might utilize CRISPR/Cas9 genome editing (Dominguez et al., 2016) in a cell type known to exhibit the *TMEM106B* eQTL effect, such as LCLs. Specifically, CRISPR/Cas9-mediated homologous recombination could be employed to mutate all three SNPs within the CTCF site to the alternative alleles, using a DNA donor template (homology directed repair) (Dominguez et al., 2016). I would mutate these SNPs in both risk haplotype homozygous and protective haplotype homozygous cell lines in order to confirm the expected changes in *TMEM106B* expression with RT-qPCR. Based on the direction of the published eQTL effects, I predict that replacing the protective SNP alleles with the risk alleles would increase *TMEM106B* expression, and vice versa. If mutating all three SNPs was sufficient to alter *TMEM106B* levels, I would perform follow-up experiments to mutate one SNP at a time, starting with rs1990620, the putative casual variant, in order to determine which variant or variants are truly causal.

While these experiments could theoretically confirm which variant(s) are causing the eQTL effect, additional experiments would add to our analysis of the mechanism. For example, one could perform CTCF ChIP-qPCR and Capture-C to confirm that editing of the causal variant(s) affects CTCF binding and long-range interactions, respectively. If the model presented in Chapter 4 is correct, such that increased CTCF binding due to the risk allele of rs1990620 facilitates long-range interactions in order to upregulate *TMEM106B*, I would expect that long-range interactions would be strengthened after editing protective allele homozygotes to risk allele homozygotes, and vice

versa. It is possible that increasing CTCF binding at the region containing rs1990620 might increase the strength of all long-range interactions, or that only specific interactions might be affected. Indeed, while CTCF sites that are located at TAD boundaries are known to be involved in TAD formation and prevention of inter-TAD long-range regulatory interactions (Dowen et al., 2014; Guo et al., 2015; Lupianez et al., 2015; Merkenschlager and Nora, 2016), the architectural role of CTCF sites located *within* TADs has not been investigated. In order to address which long-range interactions are responsible for altering *TMEM106B* levels, I would complement the Capture-C results with chromosome conformation capture (3C) coupled with qPCR (Denker and de Laat, 2016), in order to investigate specific interactions that appear to be affected by Capture-C after editing. Based on the Capture-C results described in Chapter 4, I predict that the most affected interactions would involve the *TMEM106B* promoter, the CTCF site containing rs1990620, and the enhancer. I would also perform Capture-C on unedited and edited heterozygous LCLs (which are risk or protective allele homozygotes at the causal variant(s) after editing), with the expectation that long-range interactions would be strongest in risk allele homozygotes, weakest in protective allele homozygotes, and intermediate in strength in heterozygotes. I would also use Capture-C to confirm that the allele-specific interactions seen in heterozygotes show equal interaction strength on both chromosomes after editing the causal variant(s) to homozygosity. This approach, however, might only be feasible for the *TMEM106B* promoter viewpoint (see Chapter 4), since the bait region of the CTCF site might not have sufficient heterozygous marker SNPs after editing.

In addition, if my model is correct, and the risk allele of rs1990620 (and potentially other SNPs within the CTCF site) increases *TMEM06B* expression by strengthening CTCF-mediated enhancer/promoter (E/P) interactions, I would perform CRISPR-mediated deletion of the enhancer in heterozygous cells, using paired single guide RNAs. I predict that deletion of the enhancer on the risk haplotype chromosome, which has stronger E/P interactions according to the model presented in Chapter 4, would result in a more significant decrease in *TMEM106B* levels than the same deletion on the protective allele chromosome, which has weaker E/P interactions.

Another limitation of the work presented in Chapter 4 that is worth considering is that with the exception of electromobility shift assays (EMSAs) that demonstrate allele-specific recruitment of CTCF to a probe containing rs1990620, no experiments were performed in brain-relevant cell types or tissues. Indeed, while the *TMEM106B* eQTL effect has been reported in brain in multiple studies (GTEx Consortium, 2015; Yu et al., 2015), it is unclear which cell type(s) the effect is coming from. However, several lines of evidence suggest an intriguing candidate: microglia. First, the *TMEM106B* eQTL effect has been detected in multiple white blood cell types, including LCLs, primary T cells, monocytes, and neutrophils (Fairfax et al., 2014; Peters et al., 2016; Raj et al., 2014). Second, the enhancer that may be responsible for the eQTL effect by engaging in allele-specific long-range interactions with the *TMEM106B* promoter appears to be transcriptionally active primarily in white blood cells (Andersson et al., 2014). Third, several lines of evidence discussed throughout Chapter 5 support a role of *TMEM106B* in modifying *GRN*-associated phenotypes. If *TMEM106B* does indeed influence disease risk and manifestation through effects on *GRN*, these effects may have a microglial origin, as microglia are one of the major producers and secretors of PGRN in the brain, particularly in response to injury or disease (Petkau and Leavitt, 2014). Finally, cell-based experiments have demonstrated that conditioned media from *Grn* null microglia is toxic to neurons (Martens et al., 2012). Thus, *TMEM106B* genotype and expression levels may affect PGRN in microglia, with resulting deleterious effects on neuronal health.

To definitively identify the cellular origin of the *TMEM106B* eQTL effect in brain, one might perform RNA CaptureSeq (Mercer et al., 2014) in either embryonic stem cell- or induced pluripotent stem cell-derived neurons and glia from individuals heterozygous at the *TMEM106B* haplotype, in order to quantify *TMEM106B* levels in an allele-specific manner using exonic marker SNPs. In such an approach, capture probes targeting the *TMEM106B* exons could be used to selectively enrich for *TMEM106B* in order to drastically increase sequencing read depth, thus allowing for the detection of modest allele-specific expression (ASE) differences. Any cell types exhibiting ASE could also be used in the aforementioned editing, Capture-C and 3C experiments, although current technologies to 1) produce brain-relevant cell lines from stem cells and 2) use them to perform

experiments that require many weeks of passage in cell culture may be limiting (Brennand et al., 2015). Thus, while it may not be technically feasible to perform all of these experiments at the current time, work along these lines in the future might most definitively demonstrate which variant or variants have functional effects on *TMEM106B* levels in human brain, as well as the mechanisms involved.

Third, it is possible that more than one genetic variant is affecting disease risk at the *TMEM106B* locus, through at least two possible mechanisms: first, multiple variants may underlie the eQTL effect, which would imply that Chapter 4 was not successful in identifying all the functional regulatory variants within the haplotype block; second, multiple variants might affect *TMEM106B* through distinct mechanisms, such that one variant affects *TMEM106B* mRNA levels, e.g. rs1990620, whereas another variant affects TMEM106B protein levels or function, e.g. rs3173615. Indeed, there is significant evidence that some disease-associated loci may harbor multiple "causal" variants, which may have additive or emergent effects on gene regulation, protein function, or both (Corradin et al., 2016; Edwards et al., 2013). It is worth noting, however, that if increased TMEM106B expression is truly the mechanism by which *TMEM106B* genotype influences disease risk, a more disease-relevant line of inquiry may be to investigate how changes in TMEM106B levels affect disease risk, rather than trying to understand the contributions of one or multiple variants to TMEM106B steady-state levels.

Fourth, and following from the last paragraph, it remains to be determined how small changes in TMEM106B levels affects risk for FTLD. While many groups have reported that *TMEM106B* overexpression causes lysosomal phenotypes and cell death (Brady et al., 2013; Busch et al., 2016; Chen-Plotkin et al., 2012; Schwenk et al., 2014; Stagi et al., 2014; Suzuki and Matsuoka, 2016), and that these phenotypes result from as little as ~2-fold overexpression (Chapter 4), it is unclear exactly how this happens, or which cell type in the brain these results are relevant to. Thus, after determining which brain cell type(s) exhibit ASE of *TMEM106B*, one might use the constructs described in Chapter 4 (Chen-Plotkin et al., 2012) to overexpress *TMEM106B* at three different levels, and quantify the enlarged lysosomal phenotype and cell death.

Finally, given the genetic modifier effects of *TMEM106B* in *GRN*+ FTLD-TDP (Cruchaga et al., 2011; Finch et al., 2011) and *C9orf72* FTLD-TDP (Gallagher et al., 2014) described earlier, it is of particular scientific and clinical interest to study TMEM106B function in the context of PGRN and C9orf72 function and/or loss-of-function. Indeed, while it is unlikely that a GWAS-nominated risk factor of small effect, i.e. TMEM106B in sporadic FTLD-TDP, would make for an effective therapeutic target in all patients, if the same genetic variants substantially modified penetrance or age at death in a subset of individuals at high risk for development of disease (i.e. *C9orf72* or *GRN* mutation carriers), this subset of individuals might be reasonable candidates for *TMEM106B*-based therapeutic interventions.

In summary, I conclude that genetic variants at the *TMEM106B* locus affect risk for FTLD-TDP, and modify the clinical course of *GRN*+ FTLD-TDP and *C9orf72*+ FTLD-TDP. I also conclude that this association is most likely due to one or more risk-associated genetic variants at this locus increasing steady-state *TMEM106B* mRNA levels, thereby increasing TMEM106B protein levels and impairing lysosomal function. It is possible, though perhaps not likely, that the amino acid-altering SNP rs3173615 affects disease risk independently by altering TMEM106B protein levels or function. While our results strongly suggest that allele-specific CTCF binding and long-range regulatory interactions may underlie the eQTL effect at *TMEM106B* and thereby affect disease risk, further experimentation (namely, genome-editing combined with 3C-based approaches) in brain-relevant cell lines could more definitively prove this hypothesis.

5.5 Summary

Frontotemporal lobar degeneration (FTLD) is a heterogeneous group of diseases characterized by progressive degeneration of the frontal and temporal lobes of the brain. FTLD is typically fatal within 5-8 years after symptom onset, and currently there are no disease-modifying treatments (Bang et al., 2015; Seelaar et al., 2011). A genome-wide association study (GWAS) identified common genetic variants at the *TMEM106B* locus on chromosome 7p21 that are associated with risk for the most common neuropathological subtype of FTLD, FTLD with TDP-43 pathology (FTLD-TDP) (Van Deerlin et al., 2010), and this association has been replicated by three independent groups (Finch et al., 2011; Hernandez et al., 2015; van der Zee et al., 2011). Furthermore, FTLD-TDP-associated variants have been associated with *TMEM106B* expression levels in human cell lines and brain (GTEx Consortium, 2015; Lappalainen et al., 2013; Liang et al., 2013; Stranger et al., 2012; Yu et al., 2015), and TMEM106B levels are increased in FTLD-TDP brains, compared to neurologically normal controls (Chen-Plotkin et al., 2012; Van Deerlin et al., 2010). Thus, increased TMEM106B levels may play a key role in FTLD pathogenesis.

In this work, I demonstrate that the FTLD-TDP risk genotype at *TMEM106B* affects age at disease onset and age at death in FTLD-TDP cases caused by Mendelian mutations in the *C9orf72* gene (Gallagher et al., 2014). I also show that the single most dysregulated microRNA in FTLD-TDP brains, microRNA-132, directly represses TMEM106B expression (Chen-Plotkin et al., 2012); thus, downregulation of microRNA-132 in FTLD-TDP likely contributes to the increased TMEM106B levels seen in disease. Furthermore, I perform an unbiased, comprehensive characterization of all candidate causal variants at the *TMEM106B* locus, and demonstrate that 1) the *TMEM106B* risk haplotype is associated with increased *TMEM106B* levels in human cell lines and brain, 2) incremental increases of TMEM106B levels in cells cause incremental increases in lysosomal abnormalities and cell death, and 3) a putative noncoding causal variant in complete LD with the sentinel GWAS SNP affects binding of the transcription factor CTCF and associated long-range chromatin looping interactions. These experiments suggest that I have identified the “causal variant’ responsible for the association of the 7p21 locus with FTLD-TDP. Finally, I have shown

that neurodegenerative risk SNPs are enriched in brain CTCF binding sites genome-wide, implicating a more general role of CTCF-mediated gene expression programs in risk for neurodegeneration.

In summary, this dissertation investigates the role of *TMEM106B* genotype and expression levels in the pathogenesis of FTLD. Specifically, I focus on understanding the mechanisms underlying the association of common genetic variation at the *TMEM106B* locus with FTLD by GWAS. Such a mechanistic understanding has only been accomplished for a few dozen of the thousands of variant-trait associations in the genome, and for only one of the ~200 variants associated with risk for neurodegenerative diseases. However, understanding the biological mechanisms underlying the thousands of disease-associated loci will be required in order to better understand, and eventually treat, complex human diseases.

BIBLIOGRAPHY

1000 Genomes Project Consortium, Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A. (2010). A map of human genome variation from population-scale sequencing. Nature *467,* 1061-1073.

1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491,* 56-65.

1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature *526,* 68-74.

Abraham, G., and Inouye, M. (2015). Genomic risk prediction of complex human disease and its clinical application. Curr. Opin. Genet. Dev. *33,* 10-16.

Adams, H.H., Verhaaren, B.F., Vrooman, H.A., Uitterlinden, A.G., Hofman, A., van Duijn, C.M., van der Lugt, A., Niessen, W.J., Vernooij, M.W., and Ikram, M.A. (2014). TMEM106B influences volume of left-sided temporal lobe and interhemispheric structures in the general population. Biol. Psychiatry *76,* 503-508.

Adrianto, I., Wen, F., Templeton, A., Wiley, G., King, J.B., Lessard, C.J., Bates, J.S., Hu, Y., Kelly, J.A., Kaufman, K.M*., et al.* (2011). Association of a functional variant downstream of TNFAIP3 with systemic lupus erythematosus. Nat. Genet. *43,* 253-258.

Agarwal, V., Bell, G.W., Nam, J.W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. Elife *4,* 10.7554/eLife.05005.

Albert, F.W., and Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. Nat. Rev. Genet. *16,* 197-212.

Alquezar, C., de la Encarnacion, A., Moreno, F., Lopez de Munain, A., and Martin-Requero, A. (2016). Progranulin deficiency induces overactivation of WNT5A expression via TNF-alpha/NF-kappaB pathway in peripheral cells from frontotemporal dementia-linked granulin mutation carriers. J. Psychiatry Neurosci. *41,* 225-239.

Al-Sarraj, S., King, A., Troakes, C., Smith, B., Maekawa, S., Bodi, I., Rogelj, B., Al-Chalabi, A., Hortobagyi, T., and Shaw, C.E. (2011). p62 positive, TDP-43 negative, neuronal cytoplasmic and intranuclear inclusions in the cerebellum and hippocampus define the pathology of C9orf72-linked FTLD and MND/ALS. Acta Neuropathol. *122,* 691-702.

Ameres, S.L., and Zamore, P.D. (2013). Diversifying microRNA sequence and function. Nat. Rev. Mol. Cell Biol. *14,* 475-488.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T*., et al.* (2014). An atlas of active enhancers across human cell types and tissues. Nature *507,* 455-461.

Arai, T., Hasegawa, M., Akiyama, H., Ikeda, K., Nonaka, T., Mori, H., Mann, D., Tsuchiya, K., Yoshida, M., Hashizume, Y., and Oda, T. (2006). TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Biochem. Biophys. Res. Commun. *351,* 602-611.

Arnold, M., Raffler, J., Pfeufer, A., Suhre, K., and Kastenmuller, G. (2015). SNiPA: an interactive, genetic variant-centered annotation browser. Bioinformatics *31,* 1334-1336.

Ash, P.E., Bieniek, K.F., Gendron, T.F., Caulfield, T., Lin, W.L., Dejesus-Hernandez, M., van Blitterswijk, M.M., Jansen-West, K., Paul, J.W.,3rd, Rademakers, R.*, et al.* (2013). Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. Neuron *77,* 639-646.

Atanasio, A., Decman, V., White, D., Ramos, M., Ikiz, B., Lee, H.C., Siao, C.J., Brydges, S., LaRosa, E., Bai, Y.*, et al.* (2016). C9orf72 ablation causes immune dysregulation characterized by leukocyte expansion, autoantibody production, and glomerulonephropathy in mice. Sci. Rep. *6,* 23204.

Aten, S., Hansen, K.F., Hoyt, K.R., and Obrietan, K. (2016). The miR-132/212 locus: a complex regulator of neuronal plasticity, gene expression and cognition. RNA Dis. *3,* e1375. Epub 2016 Aug 2.

Baker, M., Mackenzie, I.R., Pickering-Brown, S.M., Gass, J., Rademakers, R., Lindholm, C., Snowden, J., Adamson, J., Sadovnick, A.D., Rollinson, S.*, et al.* (2006). Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17. Nature *442,* 916-919.

Bang, J., Spina, S., and Miller, B.L. (2015). Frontotemporal dementia. Lancet *386,* 1672-1682.

Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics *21,* 263-265.

Battle, A., Khan, Z., Wang, S.H., Mitrano, A., Ford, M.J., Pritchard, J.K., and Gilad, Y. (2015). Genomic variation. Impact of regulatory variation from RNA to protein. Science *347,* 664-667.

Bauer, D.E., Kamran, S.C., Lessard, S., Xu, J., Fujiwara, Y., Lin, C., Shao, Z., Canver, M.C., Smith, E.C., Pinello, L.*, et al.* (2013). An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. Science *342,* 253-257.

Beck, J., Poulter, M., Hensman, D., Rohrer, J.D., Mahoney, C.J., Adamson, G., Campbell, T., Uphill, J., Borg, A., Fratta, P.*, et al.* (2013). Large C9orf72 hexanucleotide repeat expansions are seen in multiple neurodegenerative syndromes and are more frequent than expected in the UK population. Am. J. Hum. Genet. *92,* 345-353.

Benajiba, L., Le Ber, I., Camuzat, A., Lacoste, M., Thomas-Anterion, C., Couratier, P., Legallic, S., Salachas, F., Hannequin, D., Decousus, M.*, et al.* (2009). TARDBP mutations in motoneuron disease with frontotemporal lobar degeneration. Ann. Neurol. *65,* 470-473.

Benussi, A., Padovani, A., and Borroni, B. (2015). Phenotypic Heterogeneity of Monogenic Frontotemporal Dementia. Front. Aging Neurosci. *7,* 171.

Berndt, S.I., Gustafsson, S., Magi, R., Ganna, A., Wheeler, E., Feitosa, M.F., Justice, A.E., Monda, K.L., Croteau-Chonka, D.C., Day, F.R.*, et al.* (2013). Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. Nat. Genet. *45,* 501-512.

Bodea, L.G., Eckert, A., Ittner, L.M., Piguet, O., and Gotz, J. (2016). Tau physiology and pathomechanisms in frontotemporal lobar degeneration. J. Neurochem. *138 Suppl 1,* 71-94.

Boxer, A.L., Mackenzie, I.R., Boeve, B.F., Baker, M., Seeley, W.W., Crook, R., Feldman, H., Hsiung, G.Y., Rutherford, N., Laluz, V.*, et al.* (2011). Clinical, neuroimaging and neuropathological features of a new chromosome 9p-linked FTD-ALS family. J. Neurol. Neurosurg. Psychiatry. *82,* 196-203.

Brady, O.A., Zheng, Y., Murphy, K., Huang, M., and Hu, F. (2013). The frontotemporal lobar degeneration risk factor, TMEM106B, regulates lysosomal morphology and function. Hum. Mol. Genet. *22,* 685-695.

Brennand, K.J., Marchetto, M.C., Benvenisty, N., Brustle, O., Ebert, A., Izpisua Belmonte, J.C., Kaykas, A., Lancaster, M.A., Livesey, F.J., McConnell, M.J.*, et al.* (2015). Creating Patient-Specific Neural Cells for the In Vitro Study of Brain Disorders. Stem Cell. Reports *5,* 933-945.

Brooks, B.R., Miller, R.G., Swash, M., Munsat, T.L., and World Federation of Neurology Research Group on Motor Neuron Diseases. (2000). El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. Amyotroph Lateral Scler. Other Motor Neuron. Disord. *1,* 293-299.

Brown, C.D., Mangravite, L.M., and Engelhardt, B.E. (2013). Integrative modeling of eQTLs and cis-regulatory elements suggests mechanisms underlying cell type specificity of eQTLs. PLoS Genet. *9,* e1003649.

Brown, W.J., Constantinescu, E., and Farquhar, M.G. (1984). Redistribution of mannose-6-phosphate receptors induced by tunicamycin and chloroquine. J. Cell Biol. *99,* 320-326.

Busch, J.I., Martinez-Lage, M., Ashbridge, E., Grossman, M., Van Deerlin, V.M., Hu, F., Lee, V.M., Trojanowski, J.Q., and Chen-Plotkin, A.S. (2013). Expression of TMEM106B, the frontotemporal lobar degeneration-associated protein, in normal and diseased human brain. Acta Neuropathol. Commun. *1,* 36-5960-1-36.

Busch, J.I., Unger, T.L., Jain, N., Tyler Skrinak, R., Charan, R.A., and Chen-Plotkin, A.S. (2016). Increased expression of the frontotemporal dementia risk factor TMEM106B causes C9orf72-dependent alterations in lysosomes. Hum. Mol. Genet.

Cairns, J., Freire-Pritchett, P., Wingett, S.W., Varnai, C., Dimond, A., Plagnol, V., Zerbino, D., Schoenfelder, S., Javierre, B.M., Osborne, C., Fraser, P., and Spivakov, M. (2016). CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. Genome Biol. *17,* 127-016-0992-2.

Campbell, M.C., and Tishkoff, S.A. (2008). African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. Annu. Rev. Genomics Hum. Genet. *9,* 403-433.

Capell, A., Liebscher, S., Fellerer, K., Brouwers, N., Willem, M., Lammich, S., Gijselinck, I., Bittner, T., Carlson, A.M., Sasse, F.*, et al.* (2011). Rescue of progranulin deficiency associated with frontotemporal lobar degeneration by alkalizing reagents and inhibition of vacuolar ATPase. J. Neurosci. *31,* 1885-1894.

Chen, S., and Zheng, J.C. (2012). Translational Neurodegeneration, a platform to share knowledge and experience in translational study of neurodegenerative diseases. Transl. Neurodegener *1,* 1-9158-1-1.

Chen, Y., Li, S., Su, L., Sheng, J., Lv, W., Chen, G., and Xu, Z. (2015). Association of progranulin polymorphism rs5848 with neurodegenerative diseases: a meta-analysis. J. Neurol. *262,* 814-822.

Chen-Plotkin, A.S., Geser, F., Plotkin, J.B., Clark, C.M., Kwong, L.K., Yuan, W., Grossman, M., Van Deerlin, V.M., Trojanowski, J.Q., and Lee, V.M. (2008). Variations in the progranulin gene affect global gene expression in frontotemporal lobar degeneration. Hum. Mol. Genet. *17,* 1349-1362.

Chen-Plotkin, A.S., Hu, W.T., Siderowf, A., Weintraub, D., Goldmann Gross, R., Hurtig, H.I., Xie, S.X., Arnold, S.E., Grossman, M., Clark, C.M.*, et al.* (2011). Plasma epidermal growth factor levels predict cognitive decline in Parkinson disease. Ann. Neurol. *69,* 655-663.

Chen-Plotkin, A.S., Lee, V.M., and Trojanowski, J.Q. (2010a). TAR DNA-binding protein 43 in neurodegenerative disease. Nat. Rev. Neurol. *6,* 211-220.

Chen-Plotkin, A.S., Unger, T.L., Gallagher, M.D., Bill, E., Kwong, L.K., Volpicelli-Daley, L., Busch, J.I., Akle, S., Grossman, M., Van Deerlin, V., Trojanowski, J.Q., and Lee, V.M. (2012). TMEM106B, the risk gene for frontotemporal dementia, is regulated by the microRNA-132/212 cluster and affects progranulin pathways. J. Neurosci. *32,* 11213-11227.

Chen-Plotkin, A.S., Xiao, J., Geser, F., Martinez-Lage, M., Grossman, M., Unger, T., Wood, E.M., Van Deerlin, V.M., Trojanowski, J.Q., and Lee, V.M. (2010b). Brain progranulin expression in GRN-associated frontotemporal lobar degeneration. Acta Neuropathol. *119,* 111-122.

Chew, J., Gendron, T.F., Prudencio, M., Sasaguri, H., Zhang, Y.J., Castanedes-Casey, M., Lee, C.W., Jansen-West, K., Kurti, A., Murray, M.E.*, et al.* (2015). Neurodegeneration. C9ORF72 repeat expansions in mice cause TDP-43 pathology, neuronal loss, and behavioral deficits. Science *348,* 1151-1154.

Church, C., Moir, L., McMurray, F., Girard, C., Banks, G.T., Teboul, L., Wells, S., Bruning, J.C., Nolan, P.M., Ashcroft, F.M., and Cox, R.D. (2010). Overexpression of Fto leads to increased food intake and results in obesity. Nat. Genet. *42,* 1086-1092.

Ciura, S., Lattante, S., Le Ber, I., Latouche, M., Tostivint, H., Brice, A., and Kabashi, E. (2013). Loss of function of C9orf72 causes motor deficits in a zebrafish model of Amyotrophic Lateral Sclerosis. Ann. Neurol.

Ciura, S., Sellier, C., Campanari, M.L., Charlet-Berguerand, N., and Kabashi, E. (2016). The most prevalent genetic cause of ALS-FTD, C9orf72 synergizes the toxicity of ATXN2 intermediate polyglutamine repeats through the autophagy pathway. Autophagy *12,* 1406-1408.

Claussnitzer, M., Dankel, S.N., Kim, K.H., Quon, G., Meuleman, W., Haugen, C., Glunk, V., Sousa, I.S., Beaudry, J.L., Puviindran, V.*, et al.* (2015). FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. N. Engl. J. Med. *373,* 895-907.

Clayton, E.L., Mizielinska, S., Edgar, J.R., Nielsen, T.T., Marshall, S., Norona, F.E., Robbins, M., Damirji, H., Holm, I.E., Johannsen, P.*, et al.* (2015). Frontotemporal dementia caused by CHMP2B mutation is characterised by neuronal lysosomal storage pathology. Acta Neuropathol. *130,* 511-523.

Cohen-Kaplan, V., Livneh, I., Avni, N., Cohen-Rosenzweig, C., and Ciechanover, A. (2016). The ubiquitin-proteasome system and autophagy: Coordinated and independent activities. Int. J. Biochem. Cell Biol. *79,* 403-418.

Cooper, G.M., and Shendure, J. (2011). Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. Nat. Rev. Genet. *12,* 628-640.

Cooper-Knock, J., Walsh, M.J., Higginbottom, A., Robin Highley, J., Dickman, M.J., Edbauer, D., Ince, P.G., Wharton, S.B., Wilson, S.A., Kirby, J., Hautbergue, G.M., and Shaw, P.J. (2014). Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. Brain *137,* 2040-2051.

Corradin, O., Cohen, A.J., Luppino, J.M., Bayles, I.M., Schumacher, F.R., and Scacheri, P.C. (2016). Modeling disease risk through analysis of physical interactions between genetic variants within chromatin regulatory circuitry. Nat. Genet.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A.*, et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. U. S. A. *107,* 21931-21936.

Crick, F. (1970). Central dogma of molecular biology. Nature *227,* 561-563.

Cruchaga, C., Graff, C., Chiang, H.H., Wang, J., Hinrichs, A.L., Spiegel, N., Bertelsen, S., Mayo, K., Norton, J.B., Morris, J.C., and Goate, A. (2011). Association of TMEM106B gene polymorphism with age at onset in granulin mutation carriers and plasma granulin protein levels. Arch. Neurol. *68,* 581-586.

Cruts, M., Gijselinck, I., van der Zee, J., Engelborghs, S., Wils, H., Pirici, D., Rademakers, R., Vandenberghe, R., Dermaut, B., Martin, J.J.*, et al.* (2006). Null mutations in progranulin cause ubiquitin-positive frontotemporal dementia linked to chromosome 17q21. Nature *442,* 920-924.

Cruts, M., Gijselinck, I., Van Langenhove, T., van der Zee, J., and Van Broeckhoven, C. (2013). Current insights into the C9orf72 repeat expansion diseases of the FTLD/ALS spectrum. Trends Neurosci.

Cruts, M., and Van Broeckhoven, C. (2008). Loss of progranulin function in frontotemporal lobar degeneration. Trends Genet. *24,* 186-194.

Cutting, G.R. (2010). Modifier genes in Mendelian disorders: the example of cystic fibrosis. Ann. N. Y. Acad. Sci. *1214,* 57-69.

Davies, J.O., Telenius, J.M., McGowan, S.J., Roberts, N.A., Taylor, S., Higgs, D.R., and Hughes, J.R. (2016). Multiplexed analysis of chromosome conformation at vastly improved sensitivity. Nat. Methods *13,* 74-80.

de la Encarnacion, A., Alquezar, C., and Martin-Requero, A. (2016). Increased Wnt Signaling and Reduced Viability in a Neuronal Model of Progranulin-Deficient Frontotemporal Lobar Degeneration. Mol. Neurobiol. *53,* 7107-7118.

Degner, J.F., Pai, A.A., Pique-Regi, R., Veyrieras, J.B., Gaffney, D.J., Pickrell, J.K., De Leon, S., Michelini, K., Lewellen, N., Crawford, G.E.*, et al.* (2012). DNase I sensitivity QTLs are a major determinant of human expression variation. Nature *482,* 390-394.

DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J.*, et al.* (2011). Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron *72,* 245-256.

Denker, A., and de Laat, W. (2016). The second decade of 3C technologies: detailed insights into nuclear organization. Genes Dev. *30,* 1357-1382.

Diao, Y., Li, B., Meng, Z., Jung, I., Lee, A.Y., Dixon, J., Maliskova, L., Guan, K.L., Shen, Y., and Ren, B. (2016). A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9-mediated genetic screening. Genome Res. *26,* 397-405.

Dickel, D.E., Zhu, Y., Nord, A.S., Wylie, J.N., Akiyama, J.A., Afzal, V., Plajzer-Frick, I., Kirkpatrick, A., Gottgens, B., Bruneau, B.G., Visel, A., and Pennacchio, L.A. (2014). Function-based identification of mammalian enhancers using site-specific integration. Nat. Methods *11,* 566-571.

Dimas, A.S., Deutsch, S., Stranger, B.E., Montgomery, S.B., Borel, C., Attar-Cohen, H., Ingle, C., Beazley, C., Gutierrez Arcelus, M., Sekowska, M.*, et al.* (2009). Common regulatory variation impacts gene expression in a cell type-dependent manner. Science *325,* 1246-1250.

Ding, Z., Ni, Y., Timmer, S.W., Lee, B.K., Battenhouse, A., Louzada, S., Yang, F., Dunham, I., Crawford, G.E., Lieb, J.D.*, et al.* (2014). Quantitative genetics of CTCF binding reveal local sequence effects and different modes of X-chromosome association. PLoS Genet. *10,* e1004798.

Dixon, A.L., Liang, L., Moffatt, M.F., Chen, W., Heath, S., Wong, K.C., Taylor, J., Burnett, E., Gut, I., Farrall, M.*, et al.* (2007). A genome-wide association study of global gene expression. Nat. Genet. *39,* 1202-1207.

Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W.*, et al.* (2015). Chromatin architecture reorganization during stem cell differentiation. Nature *518,* 331-336.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485,* 376-380.

Dominguez, A.A., Lim, W.A., and Qi, L.S. (2016). Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. Nat. Rev. Mol. Cell Biol. *17,* 5-15.

Donnelly, C.J., Zhang, P.W., Pham, J.T., Haeusler, A.R., Mistry, N.A., Vidensky, S., Daley, E.L., Poth, E.M., Hoover, B., Fines, D.M.*, et al.* (2013). RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. Neuron *80,* 415-428.

Dowen, J.M., Fan, Z.P., Hnisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schuijers, J., Lee, T.I., Zhao, K., and Young, R.A. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell *159,* 374-387.

Dunning, A.M., Michailidou, K., Kuchenbaecker, K.B., Thompson, D., French, J.D., Beesley, J., Healey, C.S., Kar, S., Pooley, K.A., Lopez-Knowles, E.*, et al.* (2016). Breast cancer risk variants at 6q25 display different phenotype associations and regulate ESR1, RMND1 and CCDC170. Nat. Genet. *48,* 374-386.

Edwards, S.L., Beesley, J., French, J.D., and Dunning, A.M. (2013). Beyond GWASs: illuminating the dark road from association to function. Am. J. Hum. Genet. *93,* 779-797.

Eisele, Y.S., Monteiro, C., Fearns, C., Encalada, S.E., Wiseman, R.L., Powers, E.T., and Kelly, J.W. (2015). Targeting protein aggregation for the treatment of degenerative diseases. Nat. Rev. Drug Discov. *14,* 759-780.

ENCODE Project Consortium, Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., and Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489,* 57-74.

Engel, K.L., Mackiewicz, M., Hardigan, A.A., Myers, R.M., and Savic, D. (2016). Decoding transcriptional enhancers: Evolving from annotation to functional interpretation. Semin. Cell Dev. Biol.

Fairfax, B.P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E., Jostins, L., Plant, K., Andrews, R., McGee, C., and Knight, J.C. (2014). Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science *343,* 1246949.

FANTOM Consortium and the RIKEN PMI and CLST (DGT), Forrest, A.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J., Haberle, V., Lassmann, T., Kulakovskiy, I.V., Lizio, M.*, et al.* (2014). A promoter-level mammalian expression atlas. Nature *507,* 462-470.

Farg, M.A., Sundaramoorthy, V., Sultana, J.M., Yang, S., Atkinson, R.A., Levina, V., Halloran, M.A., Gleeson, P.A., Blair, I.P., Soo, K.Y., King, A.E., and Atkin, J.D. (2014). C9ORF72, implicated in amytrophic lateral sclerosis and frontotemporal dementia, regulates endosomal trafficking. Hum. Mol. Genet. *23,* 3579-3595.

Farh, K.K., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shoresh, N., Whitton, H., Ryan, R.J., Shishkin, A.A.*, et al.* (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature *518,* 337-343.

Ferrari, R., Hernandez, D.G., Nalls, M.A., Rohrer, J.D., Ramasamy, A., Kwok, J.B., Dobson-Stone, C., Brooks, W.S., Schofield, P.R., Halliday, G.M.*, et al.* (2014). Frontotemporal dementia and its subtypes: a genome-wide association study. Lancet Neurol. *13,* 686-699.

Filiano, A.J., Martens, L.H., Young, A.H., Warmus, B.A., Zhou, P., Diaz-Ramirez, G., Jiao, J., Zhang, Z., Huang, E.J., Gao, F.B., Farese, R.V.,Jr, and Roberson, E.D. (2013). Dissociation of

frontotemporal dementia-related deficits and neuroinflammation in progranulin haploinsufficient mice. J. Neurosci. *33,* 5352-5361.

Filipowicz, W., Bhattacharyya, S.N., and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat. Rev. Genet. *9,* 102-114.

Finch, N., Carrasquillo, M.M., Baker, M., Rutherford, N.J., Coppola, G., Dejesus-Hernandez, M., Crook, R., Hunter, T., Ghidoni, R., Benussi, L.*, et al.* (2011). TMEM106B regulates progranulin levels and the penetrance of FTLD in GRN mutation carriers. Neurology *76,* 467-474.

Fischer, J., Koch, L., Emmerling, C., Vierkotten, J., Peters, T., Bruning, J.C., and Ruther, U. (2009). Inactivation of the Fto gene protects from obesity. Nature *458,* 894-898.

Fogarty, M.P., Cannon, M.E., Vadlamudi, S., Gaulton, K.J., and Mohlke, K.L. (2014). Identification of a regulatory variant that binds FOXA1 and FOXA2 at the CDC123/CAMK1D type 2 diabetes GWAS locus. PLoS Genet. *10,* e1004633.

Fratta, P., Mizielinska, S., Nicoll, A.J., Zloh, M., Fisher, E.M., Parkinson, G., and Isaacs, A.M. (2012). C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. Sci. Rep. *2,* 1016.

Fratta, P., Poulter, M., Lashley, T., Rohrer, J.D., Polke, J.M., Beck, J., Ryan, N., Hensman, D., Mizielinska, S., Waite, A.J.*, et al.* (2013). Homozygosity for the C9orf72 GGGGCC repeat expansion in frontotemporal dementia. Acta Neuropathol. *126,* 401-409.

Freibaum, B.D., Lu, Y., Lopez-Gonzalez, R., Kim, N.C., Almeida, S., Lee, K.H., Badders, N., Valentine, M., Miller, B.L., Wong, P.C.*, et al.* (2015). GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. Nature *525,* 129-133.

Fu, J., Wolfs, M.G., Deelen, P., Westra, H.J., Fehrmann, R.S., Te Meerman, G.J., Buurman, W.A., Rensen, S.S., Groen, H.J., Weersma, R.K.*, et al.* (2012). Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. PLoS Genet. *8,* e1002431.

Fulco, C.P., Munschauer, M., Anyoha, R., Munson, G., Grossman, S.R., Perez, E.M., Kane, M., Cleary, B., Lander, E.S., and Engreitz, J.M. (2016). Systematic mapping of functional enhancer-promoter connections with CRISPR interference. Science *354,* 769-773.

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M.*, et al.* (2002). The structure of haplotype blocks in the human genome. Science *296,* 2225-2229.

Gaj, T., Sirk, S.J., Shui, S.L., and Liu, J. (2016). Genome-Editing Technologies: Principles and Applications. Cold Spring Harb Perspect. Biol. *8,* 10.1101/cshperspect.a023754.

Gallagher, M.D., Suh, E., Grossman, M., Elman, L., McCluskey, L., Van Swieten, J.C., Al-Sarraj, S., Neumann, M., Gelpi, E., Ghetti, B.*, et al.* (2014). TMEM106B is a genetic modifier of frontotemporal lobar degeneration with C9orf72 hexanucleotide repeat expansions. Acta Neuropathol. *127,* 407-418.

Gao, X., Becker, L.C., Becker, D.M., Starmer, J.D., and Province, M.A. (2010a). Avoiding the high Bonferroni penalty in genome-wide association studies. Genet. Epidemiol. *34,* 100-105.

Gao, X., Shin, Y.H., Li, M., Wang, F., Tong, Q., and Zhang, P. (2010b). The fat mass and obesity associated gene FTO functions in the brain to regulate postnatal growth in mice. PLoS One *5,* e14005.

Gass, J., Cannon, A., Mackenzie, I.R., Boeve, B., Baker, M., Adamson, J., Crook, R., Melquist, S., Kuntz, K., Petersen, R.*, et al.* (2006). Mutations in progranulin are a major cause of ubiquitin-positive frontotemporal lobar degeneration. Hum. Mol. Genet. *15,* 2988-3001.

Gendron, T.F., Bieniek, K.F., Zhang, Y.J., Jansen-West, K., Ash, P.E., Caulfield, T., Daughrity, L., Dunmore, J.H., Castanedes-Casey, M., Chew, J.*, et al.* (2013). Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. Acta Neuropathol. *126,* 829-844.

Gerstein, M.B., Kundaje, A., Hariharan, M., Landt, S.G., Yan, K.K., Cheng, C., Mu, X.J., Khurana, E., Rozowsky, J., Alexander, R.*, et al.* (2012). Architecture of the human regulatory network derived from ENCODE data. Nature *489,* 91-100.

Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., and Furlong, E.E. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. Nature *512,* 96-100.

Ghosh, S., and Collins, F.S. (1996). The geneticist's approach to complex disease. Annu. Rev. Med. *47,* 333-353.

Ghoussaini, M., Edwards, S.L., Michailidou, K., Nord, S., Cowper-Sal Lari, R., Desai, K., Kar, S., Hillman, K.M., Kaufmann, S., Glubb, D.M.*, et al.* (2014). Evidence that breast cancer risk at the 2q35 locus is mediated through IGFBP5 regulation. Nat. Commun. *4,* 4999.

Gibson, G. (2012). Rare and common variants: twenty arguments. Nat. Rev. Genet. *13,* 135-145.

Gijselinck, I., Van Langenhove, T., van der Zee, J., Sleegers, K., Philtjens, S., Kleinberger, G., Janssens, J., Bettens, K., Van Cauwenberghe, C., Pereson, S.*, et al.* (2012). A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: a gene identification study. Lancet Neurol. *11,* 54-65.

Gitcho, M.A., Baloh, R.H., Chakraverty, S., Mayo, K., Norton, J.B., Levitch, D., Hatanpaa, K.J., White, C.L.,3rd, Bigio, E.H., Caselli, R.*, et al.* (2008). TDP-43 A315T mutation in familial motor neuron disease. Ann. Neurol. *63,* 535-538.

Glubb, D.M., Maranian, M.J., Michailidou, K., Pooley, K.A., Meyer, K.B., Kar, S., Carlebur, S., O'Reilly, M., Betts, J.A., Hillman, K.M.*, et al.* (2015). Fine-scale mapping of the 5q11.2 breast cancer locus reveals at least three independent risk variants regulating MAP3K1. Am. J. Hum. Genet. *96,* 5-20.

Goodall, E.F., Heath, P.R., Bandmann, O., Kirby, J., and Shaw, P.J. (2013). Neuronal dark matter: the emerging role of microRNAs in neurodegeneration. Front. Cell. Neurosci. *7,* 178.

Gorno-Tempini, M.L., Hillis, A.E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S.F., Ogar, J.M., Rohrer, J.D., Black, S., Boeve, B.F.*, et al.* (2011). Classification of primary progressive aphasia and its variants. Neurology *76,* 1006-1014.

GTEx Consortium. (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science *348,* 648-660.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y.*, et al.* (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell *162,* 900-910.

Guthridge, J.M., Lu, R., Sun, H., Sun, C., Wiley, G.B., Dominguez, N., Macwana, S.R., Lessard, C.J., Kim-Howard, X., Cobb, B.L.*, et al.* (2014). Two functional lupus-associated BLK promoter variants control cell-type- and developmental-stage-specific transcription. Am. J. Hum. Genet. *94,* 586-598.

Haeusler, A.R., Donnelly, C.J., Periz, G., Simko, E.A., Shaw, P.G., Kim, M.S., Maragakis, N.J., Troncoso, J.C., Pandey, A., Sattler, R., Rothstein, J.D., and Wang, J. (2014). C9orf72 nucleotide repeat structures initiate molecular cascades of disease. Nature *507,* 195-200.

Hannon, E., Spiers, H., Viana, J., Pidsley, R., Burrage, J., Murphy, T.M., Troakes, C., Turecki, G., O'Donovan, M.C., Schalkwyk, L.C., Bray, N.J., and Mill, J. (2016). Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. Nat. Neurosci. *19,* 48-54.

Hansen, K.F., Sakamoto, K., Wayman, G.A., Impey, S., and Obrietan, K. (2010). Transgenic miR132 alters neuronal spine density and impairs novel object recognition memory. PLoS One *5,* e15497.

Harismendy, O., Notani, D., Song, X., Rahim, N.G., Tanasa, B., Heintzman, N., Ren, B., Fu, X.D., Topol, E.J., Rosenfeld, M.G., and Frazer, K.A. (2011). 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. Nature *470,* 264-268.

Harms, M., Benitez, B.A., Cairns, N., Cooper, B., Cooper, P., Mayo, K., Carrell, D., Faber, K., Williamson, J., Bird, T.*, et al.* (2013). C9orf72 hexanucleotide repeat expansions in clinical Alzheimer disease. JAMA Neurol. *70,* 736-741.

Hause, R.J., Stark, A.L., Antao, N.N., Gorsic, L.K., Chung, S.H., Brown, C.D., Wong, S.S., Gill, D.F., Myers, J.L., To, L.A.*, et al.* (2014). Identification and validation of genetic variants that influence transcription factor and cell signaling protein levels. Am. J. Hum. Genet. *95,* 194-208.

He, Z., and Bateman, A. (2003). Progranulin (granulin-epithelin precursor, PC-cell-derived growth factor, acrogranin) mediates tissue repair and tumorigenesis. J. Mol. Med. (Berl) *81,* 600-612.

Hebert, S.S., Horre, K., Nicolai, L., Papadopoulou, A.S., Mandemakers, W., Silahtaroglu, A.N., Kauppinen, S., Delacourte, A., and De Strooper, B. (2008). Loss of microRNA cluster miR-29a/b-1 in sporadic Alzheimer's disease correlates with increased BACE1/beta-secretase expression. Proc. Natl. Acad. Sci. U. S. A. *105,* 6415-6420.

Heinemeyer, T., Wingender, E., Reuter, I., Hermjakob, H., Kel, A.E., Kel, O.V., Ignatieva, E.V., Ananko, E.A., Podkolodnaya, O.A., Kolpakov, F.A., Podkolodny, N.L., and Kolchanov, N.A. (1998). Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. Nucleic Acids Res. *26,* 362-367.

Hellman, L.M., and Fried, M.G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. Nat. Protoc. *2,* 1849-1861.

Hernandez, I., Rosende-Roca, M., Alegret, M., Mauleon, A., Espinosa, A., Vargas, L., Sotolongo-Grau, O., Tarraga, L., Boada, M., and Ruiz, A. (2015). Association of TMEM106B rs1990622 marker and frontotemporal dementia: evidence for a recessive effect and meta-analysis. J. Alzheimers Dis. *43,* 325-334.

Hinrichs, A.S., Raney, B.J., Speir, M.L., Rhead, B., Casper, J., Karolchik, D., Kuhn, R.M., Rosenbloom, K.R., Zweig, A.S., Haussler, D., and Kent, W.J. (2016). UCSC Data Integrator and Variant Annotation Integrator. Bioinformatics

Hirschhorn, J.N. (2005). Genetic approaches to studying common diseases and complex traits. Pediatr. Res. *57,* 74R-77R.

Hodges, J.R., Davies, R., Xuereb, J., Kril, J., and Halliday, G. (2003). Survival in frontotemporal dementia. Neurology *61,* 349-354.

Hollander, J.A., Im, H.I., Amelio, A.L., Kocerha, J., Bali, P., Lu, Q., Willoughby, D., Wahlestedt, C., Conkright, M.D., and Kenny, P.J. (2010). Striatal microRNA controls cocaine intake through CREB signalling. Nature *466,* 197-202.

Holwerda, S.J., van de Werken, H.J., Ribeiro de Almeida, C., Bergen, I.M., de Bruijn, M.J., Verstegen, M.J., Simonis, M., Splinter, E., Wijchers, P.J., Hendriks, R.W., and de Laat, W. (2013). Allelic exclusion of the immunoglobulin heavy chain locus is independent of its nuclear localization in mature B cells. Nucleic Acids Res. *41,* 6905-6916.

Hu, F., Padukkavidana, T., Vaegter, C.B., Brady, O.A., Zheng, Y., Mackenzie, I.R., Feldman, H.H., Nykjaer, A., and Strittmatter, S.M. (2010). Sortilin-mediated endocytosis determines levels of the frontotemporal dementia protein, progranulin. Neuron *68,* 654-667.

Huang, Q., Whitington, T., Gao, P., Lindberg, J.F., Yang, Y., Sun, J., Vaisanen, M.R., Szulkin, R., Annala, M., Yan, J.*, et al.* (2014). A prostate cancer susceptibility allele at 6q22 increases RFX6 expression by modulating HOXB13 chromatin binding. Nat. Genet. *46,* 126-135.

Hughes, J.R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R., and Higgs, D.R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. Nat. Genet. *46,* 205-212.

Hutton, M., Lendon, C.L., Rizzu, P., Baker, M., Froelich, S., Houlden, H., Pickering-Brown, S., Chakraverty, S., Isaacs, A., Grover, A.*, et al.* (1998). Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. Nature *393,* 702-705.

Impey, S., McCorkle, S.R., Cha-Molstad, H., Dwyer, J.M., Yochum, G.S., Boss, J.M., McWeeney, S., Dunn, J.J., Mandel, G., and Goodman, R.H. (2004). Defining the CREB regulon: a genome-wide analysis of transcription factor regulatory regions. Cell *119,* 1041-1054.

Inoue, F., and Ahituv, N. (2015). Decoding enhancers using massively parallel reporter assays. Genomics *106,* 159-164.

International HapMap 3 Consortium, Altshuler, D.M., Gibbs, R.A., Peltonen, L., Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F.*, et al.* (2010). Integrating common and rare genetic variation in diverse human populations. Nature *467,* 52-58.

Isaacs, A.M., Johannsen, P., Holm, I., Nielsen, J.E., and FReJA consortium. (2011). Frontotemporal dementia caused by CHMP2B mutations. Curr. Alzheimer Res. *8,* 246-251.

Ishikura, N., Clever, J.L., Bouzamondo-Bernstein, E., Samayoa, E., Prusiner, S.B., Huang, E.J., and DeArmond, S.J. (2005). Notch-1 activation and dendritic atrophy in prion disease. Proc. Natl. Acad. Sci. U. S. A. *102,* 886-891.

Jendrzejewski, J., He, H., Radomska, H.S., Li, W., Tomsic, J., Liyanarachchi, S., Davuluri, R.V., Nagy, R., and de la Chapelle, A. (2012). The polymorphism rs944289 predisposes to papillary thyroid carcinoma through a large intergenic noncoding RNA gene of tumor suppressor type. Proc. Natl. Acad. Sci. U. S. A. *109,* 8646-8651.

Jiang, J., Zhu, Q., Gendron, T.F., Saberi, S., McAlonis-Downes, M., Seelman, A., Stauffer, J.E., Jafar-Nejad, P., Drenner, K., Schulte, D.*, et al.* (2016). Gain of Toxicity from ALS/FTD-Linked Repeat Expansions in C9ORF72 Is Alleviated by Antisense Oligonucleotides Targeting GGGGCC-Containing RNAs. Neuron *90,* 535-550.

Johnson, G.C., and Todd, J.A. (2000). Strategies in complex disease mapping. Curr. Opin. Genet. Dev. *10,* 330-334.

Johnson, R.C., Nelson, G.W., Troyer, J.L., Lautenberger, J.A., Kessing, B.D., Winkler, C.A., and O'Brien, S.J. (2010). Accounting for multiple comparisons in a genome-wide association study (GWAS). BMC Genomics *11,* 724-2164-11-724.

Jonas, S., and Izaurralde, E. (2015). Towards a molecular understanding of microRNA-mediated gene silencing. Nat. Rev. Genet. *16,* 421-433.

Jovicic, A., Mertens, J., Boeynaems, S., Bogaert, E., Chai, N., Yamada, S.B., Paul, J.W.,3rd, Sun, S., Herdy, J.R., Bieri, G.*, et al.* (2015). Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. Nat. Neurosci. *18,* 1226-1229.

Jun, M.H., Han, J.H., Lee, Y.K., Jang, D.J., Kaang, B.K., and Lee, J.A. (2015). TMEM106B, a frontotemporal lobar dementia (FTLD) modifier, associates with FTD-3-linked CHMP2B, a complex of ESCRT-III. Mol. Brain *8,* 85-015-0177-z.

Jung, J., Nayak, A., Schaeffer, V., Starzetz, T., Kirsch, A.K., Muller, S., Dikic, I., Mittelbronn, M., and Behrends, C. (2017). Multiplex image-based autophagy RNAi screening identifies SMCR8 as ULK1 kinase activity and gene expression regulator. Elife *6,* 10.7554/eLife.23063.

Kalia, L.V., and Lang, A.E. (2015). Parkinson's disease. Lancet *386,* 896-912.

Kaplow, I.M., MacIsaac, J.L., Mah, S.M., McEwen, L.M., Kobor, M.S., and Fraser, H.B. (2015). A pooling-based approach to mapping genetic variants associated with DNA methylation. Genome Res. *25,* 907-917.

Kapoor, A., Sekar, R.B., Hansen, N.F., Fox-Talbot, K., Morley, M., Pihur, V., Chatterjee, S., Brandimarto, J., Moravec, C.S., Pulit, S.L.*, et al.* (2014). An enhancer polymorphism at the cardiomyocyte intercalated disc protein NOS1AP locus is a major regulator of the QT interval. Am. J. Hum. Genet. *94,* 854-869.

Katsnelson, A., De Strooper, B., and Zoghbi, H.Y. (2016). Neurodegeneration: From cellular concepts to clinical applications. Sci. Transl. Med. *8,* 364ps18.

Kern, A.D., and Kondrashov, F.A. (2004). Mechanisms and convergence of compensatory evolution in mammalian mitochondrial tRNAs. Nat. Genet. *36,* 1207-1212.

Kiernan, M.C., Vucic, S., Cheah, B.C., Turner, M.R., Eisen, A., Hardiman, O., Burrell, J.R., and Zoing, M.C. (2011). Amyotrophic lateral sclerosis. Lancet *377,* 942-955.

Kim, J., Inoue, K., Ishii, J., Vanti, W.B., Voronov, S.V., Murchison, E., Hannon, G., and Abeliovich, A. (2007a). A MicroRNA feedback circuit in midbrain dopamine neurons. Science *317,* 1220-1224.

Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkov, V.V., and Ren, B. (2007b). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell *128,* 1231-1245.

Kimonis, V.E., Mehta, S.G., Fulchiero, E.C., Thomasova, D., Pasquali, M., Boycott, K., Neilan, E.G., Kartashov, A., Forman, M.S., Tucker, S.*, et al.* (2008). Clinical studies in familial VCP myopathy associated with Paget disease of bone and frontotemporal dementia. Am. J. Med. Genet. A. *146A,* 745-757.

Klein, R.J., Zeiss, C., Chew, E.Y., Tsai, J.Y., Sackler, R.S., Haynes, C., Henning, A.K., SanGiovanni, J.P., Mane, S.M., Mayne, S.T.*, et al.* (2005). Complement factor H polymorphism in age-related macular degeneration. Science *308,* 385-389.

Kleinberger, G., Capell, A., Haass, C., and Van Broeckhoven, C. (2013). Mechanisms of granulin deficiency: lessons from cellular and animal models. Mol. Neurobiol. *47,* 337-360.

Knopman, D.S., and Roberts, R.O. (2011). Estimating the number of persons with frontotemporal lobar degeneration in the US population. J. Mol. Neurosci. *45,* 330-335.

Kollmann, K., Mutenda, K.E., Balleininger, M., Eckermann, E., von Figura, K., Schmidt, B., and Lubke, T. (2005). Identification of novel lysosomal matrix proteins by proteome analysis. Proteomics *5,* 3966-3978.

Komor, A.C., Badran, A.H., and Liu, D.R. (2016). CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. Cell

Kondrashov, A.S., Sunyaev, S., and Kondrashov, F.A. (2002). Dobzhansky-Muller incompatibilities in protein evolution. Proc. Natl. Acad. Sci. U. S. A. *99,* 14878-14883.

Koppers, M., Blokhuis, A.M., Westeneng, H.J., Terpstra, M.L., Zundel, C.A., Vieira de Sa, R., Schellevis, R.D., Waite, A.J., Blake, D.J., Veldink, J.H., van den Berg, L.H., and Pasterkamp, R.J. (2015). C9orf72 ablation in mice does not cause motor neuron degeneration or motor deficits. Ann. Neurol. *78,* 426-438.

Korkmaz, G., Lopes, R., Ugalde, A.P., Nevedomskaya, E., Han, R., Myacheva, K., Zwart, W., Elkon, R., and Agami, R. (2016). Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. Nat. Biotechnol. *34,* 192-198.

Kwon, I., Xiang, S., Kato, M., Wu, L., Theodoropoulos, P., Wang, T., Kim, J., Yun, J., Xie, Y., and McKnight, S.L. (2014). Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. Science *345,* 1139-1145.

Lagier-Tourenne, C., Baughn, M., Rigo, F., Sun, S., Liu, P., Li, H.R., Jiang, J., Watt, A.T., Chun, S., Katz, M.*, et al.* (2013). Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. Proc. Natl. Acad. Sci. U. S. A. *110,* E4530-9.

Lamb, C.A., Dooley, H.C., and Tooze, S.A. (2013). Endocytosis and autophagy: Shared machinery for degradation. Bioessays *35,* 34-45.

Lambert, M.A., Bickel, H., Prince, M., Fratiglioni, L., Von Strauss, E., Frydecka, D., Kiejna, A., Georges, J., and Reynish, E.L. (2014). Estimating the burden of early onset dementia; systematic review of disease prevalence. Eur. J. Neurol. *21,* 563-569.

Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P.*, et al.* (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. *22,* 1813-1831.

Lang, C.M., Fellerer, K., Schwenk, B.M., Kuhn, P.H., Kremmer, E., Edbauer, D., Capell, A., and Haass, C. (2012). Membrane orientation and subcellular localization of transmembrane protein 106B (TMEM106B), a major risk factor for frontotemporal lobar degeneration. J. Biol. Chem. *287,* 19355-19365.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10,* R25-2009-10-3-r25. Epub 2009 Mar 4.

Lappalainen, T., Sammeth, M., Friedlander, M.R., 't Hoen, P.A., Monlong, J., Rivas, M.A., Gonzalez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G.*, et al.* (2013). Transcriptome and genome sequencing uncovers functional variation in humans. Nature *501,* 506-511.

Lee, M.J., Chen, T.F., Cheng, T.W., and Chiu, M.J. (2011). rs5848 variant of progranulin gene is a risk of Alzheimer's disease in the Taiwanese population. Neurodegener Dis. *8,* 216-220.

Lee, Y., Samaco, R.C., Gatchel, J.R., Thaller, C., Orr, H.T., and Zoghbi, H.Y. (2008). miR-19, miR-101 and miR-130 co-regulate ATXN1 levels to potentially modulate SCA1 pathogenesis. Nat. Neurosci. *11,* 1137-1139.

Lee, Y.B., Chen, H.J., Peres, J.N., Gomez-Deza, J., Attig, J., Stalekar, M., Troakes, C., Nishimura, A.L., Scotter, E.L., Vance, C.*, et al.* (2013). Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. Cell. Rep. *5,* 1178-1186.

Lelli, K.M., Slattery, M., and Mann, R.S. (2012). Disentangling the many layers of eukaryotic transcriptional regulation. Annu. Rev. Genet. *46,* 43-68.

Lettre, G., and Bauer, D.E. (2016). Fetal haemoglobin in sickle-cell disease: from genetic epidemiology to new therapeutic strategies. Lancet *387,* 2554-2564.

Levine, T.P., Daniels, R.D., Gatta, A.T., Wong, L.H., and Hayes, M.J. (2013). The product of C9orf72, a gene strongly implicated in neurodegeneration, is structurally related to DENN Rab-GEFs. Bioinformatics *29,* 499-503.

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. Cell *115,* 787-798.

Li, W., Notani, D., and Rosenfeld, M.G. (2016). Enhancers as non-coding RNA transcription units: recent insights and future perspectives. Nat. Rev. Genet. *17,* 207-223.

Liang, L., Morar, N., Dixon, A.L., Lathrop, G.M., Abecasis, G.R., Moffatt, M.F., and Cookson, W.O. (2013). A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. Genome Res. *23,* 716-726.

Lippa, C.F., Rosso, A.L., Stutzbach, L.D., Neumann, M., Lee, V.M., and Trojanowski, J.Q. (2009). Transactive response DNA-binding protein 43 burden in familial Alzheimer disease and Down syndrome. Arch. Neurol. *66,* 1483-1488.

Litvan, I., Agid, Y., Calne, D., Campbell, G., Dubois, B., Duvoisin, R.C., Goetz, C.G., Golbe, L.I., Grafman, J., Growdon, J.H.*, et al.* (1996). Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): report of the NINDS-SPSP international workshop. Neurology *47,* 1-9.

Liu, E.Y., Russ, J., Wu, K., Neal, D., Suh, E., McNally, A.G., Irwin, D.J., Van Deerlin, V.M., and Lee, E.B. (2014). C9orf72 hypermethylation protects against repeat expansion-associated pathology in ALS/FTD. Acta Neuropathol. *128,* 525-541.

Liu, Y., Pattamatta, A., Zu, T., Reid, T., Bardhi, O., Borchelt, D.R., Yachnis, A.T., and Ranum, L.P. (2016). C9orf72 BAC Mouse Model with Motor Deficits and Neurodegenerative Features of ALS/FTD. Neuron *90,* 521-534.

Lui, H., Zhang, J., Makinson, S.R., Cahill, M.K., Kelley, K.W., Huang, H.Y., Shang, Y., Oldham, M.C., Martens, L.H., Gao, F.*, et al.* (2016). Progranulin Deficiency Promotes Circuit-Specific Synaptic Pruning by Microglia via Complement Activation. Cell *165,* 921-935.

Lupianez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R.*, et al.* (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161,* 1012-1025.

Ma, Y., Matsuwaki, T., Yamanouchi, K., and Nishihara, M. (2016). Progranulin Protects Hippocampal Neurogenesis via Suppression of Neuroinflammatory Responses Under Acute Immune Stress. Mol. Neurobiol.

Mackenzie, I.R., and Neumann, M. (2016). Molecular neuropathology of frontotemporal dementia: insights into disease mechanisms from postmortem studies. J. Neurochem. *138 Suppl 1,* 54-70.

Mackenzie, I.R., Neumann, M., Baborie, A., Sampathu, D.M., Du Plessis, D., Jaros, E., Perry, R.H., Trojanowski, J.Q., Mann, D.M., and Lee, V.M. (2011). A harmonized classification system for FTLD-TDP pathology. Acta Neuropathol. *122,* 111-113.

Magill, S.T., Cambronne, X.A., Luikart, B.W., Lioy, D.T., Leighton, B.H., Westbrook, G.L., Mandel, G., and Goodman, R.H. (2010). microRNA-132 regulates dendritic growth and arborization of newborn neurons in the adult hippocampus. Proc. Natl. Acad. Sci. U. S. A. *107,* 20382-20387.

Majounie, E., Renton, A.E., Mok, K., Dopper, E.G., Waite, A., Rollinson, S., Chio, A., Restagno, G., Nicolaou, N., Simon-Sanchez, J.*, et al.* (2012). Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. Lancet Neurol. *11,* 323-330.

Mann, D.M., Rollinson, S., Robinson, A., Bennion Callister, J., Thompson, J.C., Snowden, J.S., Gendron, T., Petrucelli, L., Masuda-Suzukake, M., Hasegawa, M., Davidson, Y., and Pickering-Brown, S. (2013). Dipeptide repeat proteins are present in the p62 positive inclusions in patients with frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9ORF72. Acta Neuropathol. Commun. *1,* 68-5960-1-68.

Martens, L.H., Zhang, J., Barmada, S.J., Zhou, P., Kamiya, S., Sun, B., Min, S.W., Gan, L., Finkbeiner, S., Huang, E.J., and Farese, R.V.,Jr. (2012). Progranulin deficiency promotes neuroinflammation and neuron loss following toxin-induced injury. J. Clin. Invest. *122,* 3955-3959.

Martini-Stoica, H., Xu, Y., Ballabio, A., and Zheng, H. (2016). The Autophagy-Lysosomal Pathway in Neurodegeneration: A TFEB Perspective. Trends Neurosci. *39,* 221-234.

Mathelier, A., Zhao, X., Zhang, A.W., Parcy, F., Worsley-Hunt, R., Arenillas, D.J., Buchman, S., Chen, C.Y., Chou, A., Ienasescu, H.*, et al.* (2014). JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Res. *42,* D142-7.

Maurano, M.T., Haugen, E., Sandstrom, R., Vierstra, J., Shafer, A., Kaul, R., and Stamatoyannopoulos, J.A. (2015). Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo. Nat. Genet. *47,* 1393-1401.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J.*, et al.* (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337,* 1190-1195.

McKhann, G.M., Albert, M.S., Grossman, M., Miller, B., Dickson, D., Trojanowski, J.Q., and Work Group on Frontotemporal Dementia and Pick's Disease. (2001). Clinical and pathological diagnosis of frontotemporal dementia: report of the Work Group on Frontotemporal Dementia and Pick's Disease. Arch. Neurol. *58,* 1803-1809.

McMillan, C.T., Russ, J., Wood, E.M., Irwin, D.J., Grossman, M., McCluskey, L., Elman, L., Van Deerlin, V., and Lee, E.B. (2015). C9orf72 promoter hypermethylation is neuroprotective: Neuroimaging and neuropathologic evidence. Neurology *84,* 1622-1630.

Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., Feizi, S., Gnirke, A., Callan, C.G.,Jr, Kinney, J.B.*, et al.* (2012). Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat. Biotechnol. *30,* 271-277.

Mendenhall, E.M., Williamson, K.E., Reyon, D., Zou, J.Y., Ram, O., Joung, J.K., and Bernstein, B.E. (2013). Locus-specific editing of histone modifications at endogenous enhancers. Nat. Biotechnol. *31,* 1133-1136.

Mercer, T.R., Clark, M.B., Crawford, J., Brunck, M.E., Gerhardt, D.J., Taft, R.J., Nielsen, L.K., Dinger, M.E., and Mattick, J.S. (2014). Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. Nat. Protoc. *9,* 989-1009.

Merkenschlager, M., and Nora, E.P. (2016). CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. Annu. Rev. Genomics Hum. Genet.

Minami, S.S., Min, S.W., Krabbe, G., Wang, C., Zhou, Y., Asgarov, R., Li, Y., Martens, L.H., Elia, L.P., Ward, M.E*., et al.* (2014). Progranulin protects against amyloid beta deposition and toxicity in Alzheimer's disease mouse models. Nat. Med. *20,* 1157-1164.

Mizielinska, S., Gronke, S., Niccoli, T., Ridler, C.E., Clayton, E.L., Devoy, A., Moens, T., Norona, F.E., Woollacott, I.O., Pietrzyk, J*., et al.* (2014). C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. Science *345,* 1192-1194.

Mizielinska, S., Lashley, T., Norona, F.E., Clayton, E.L., Ridler, C.E., Fratta, P., and Isaacs, A.M. (2013). C9orf72 frontotemporal lobar degeneration is characterised by frequent neuronal sense and antisense RNA foci. Acta Neuropathol. *126,* 845-857.

Mori, K., Weng, S.M., Arzberger, T., May, S., Rentzsch, K., Kremmer, E., Schmid, B., Kretzschmar, H.A., Cruts, M., Van Broeckhoven, C., Haass, C., and Edbauer, D. (2013). The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. Science *339,* 1335-1338.

Mukherjee, O., Wang, J., Gitcho, M., Chakraverty, S., Taylor-Reinwald, L., Shears, S., Kauwe, J.S., Norton, J., Levitch, D., Bigio, E.H*., et al.* (2008). Molecular characterization of novel progranulin (GRN) mutations in frontotemporal dementia. Hum. Mutat. *29,* 512-521.

Mukhopadhyay, A., Funato, K., and Stahl, P.D. (1997). Rab7 regulates transport from early to late endocytic compartments in Xenopus oocytes. J. Biol. Chem. *272,* 13055-13059.

Murray, M.E., DeJesus-Hernandez, M., Rutherford, N.J., Baker, M., Duara, R., Graff-Radford, N.R., Wszolek, Z.K., Ferman, T.J., Josephs, K.A., Boylan, K.B., Rademakers, R., and Dickson, D.W. (2011). Clinical and neuropathologic heterogeneity of c9FTD/ALS associated with hexanucleotide repeat expansion in C9ORF72. Acta Neuropathol. *122,* 673-690.

Musunuru, K., Strong, A., Frank-Kamenetsky, M., Lee, N.E., Ahfeldt, T., Sachs, K.V., Li, X., Li, H., Kuperwasser, N., Ruda, V.M*., et al.* (2010). From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature *466,* 714-719.

Nakaoka, H., Gurumurthy, A., Hayano, T., Ahmadloo, S., Omer, W.H., Yoshihara, K., Yamamoto, A., Kurose, K., Enomoto, T., Akira, S., Hosomichi, K., and Inoue, I. (2016). Allelic Imbalance in Regulation of ANRIL through Chromatin Interaction at 9p21 Endometriosis Risk Locus. PLoS Genet. *12,* e1005893.

Nalls, M.A., Pankratz, N., Lill, C.M., Do, C.B., Hernandez, D.G., Saad, M., DeStefano, A.L., Kara, E., Bras, J., Sharma, M*., et al.* (2014). Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nat. Genet. *46,* 989-993.

Nelson, P.T., Wang, W.X., Partch, A.B., Monsell, S.E., Valladares, O., Ellingson, S.R., Wilfred, B.R., Naj, A.C., Wang, L.S., Kukull, W.A., and Fardo, D.W. (2015). Reassessment of risk genotypes

(GRN, TMEM106B, and ABCC9 variants) associated with hippocampal sclerosis of aging pathology. J. Neuropathol. Exp. Neurol. *74,* 75-84.

Neumann, M., Sampathu, D.M., Kwong, L.K., Truax, A.C., Micsenyi, M.C., Chou, T.T., Bruce, J., Schuck, T., Grossman, M., Clark, C.M.*, et al.* (2006). Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science *314,* 130-133.

Nicholson, A.M., Finch, N.A., Wojtas, A., Baker, M.C., Perkerson, R.B., Castanedes-Casey, M., Rousseau, L., Benussi, L., Binetti, G., Ghidoni, R.*, et al.* (2013). TMEM106B p.T185S regulates TMEM106B protein levels: implications for frontotemporal dementia. J. Neurochem.

Nicolae, D.L., Gamazon, E., Zhang, W., Duan, S., Dolan, M.E., and Cox, N.J. (2010). Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet. *6,* e1000888.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J.*, et al.* (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485,* 381-385.

Ong, C.T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. Nat. Rev. Genet. *15,* 234-246.

O'Rourke, J.G., Bogdanik, L., Muhammad, A.K., Gendron, T.F., Kim, K.J., Austin, A., Cady, J., Liu, E.Y., Zarrow, J., Grant, S.*, et al.* (2015). C9orf72 BAC Transgenic Mice Display Typical Pathologic Features of ALS/FTD. Neuron *88,* 892-901.

O'Rourke, J.G., Bogdanik, L., Yanez, A., Lall, D., Wolf, A.J., Muhammad, A.K., Ho, R., Carmona, S., Vit, J.P., Zarrow, J.*, et al.* (2016). C9orf72 is required for proper macrophage and microglial function in mice. Science *351,* 1324-1329.

Pai, A.A., Pritchard, J.K., and Gilad, Y. (2015). The genetic and mechanistic basis for variation in gene regulation. PLoS Genet. *11,* e1004857.

Painter, J.N., Kaufmann, S., O'Mara, T.A., Hillman, K.M., Sivakumaran, H., Darabi, H., Cheng, T.H., Pearson, J., Kazakoff, S., Waddell, N.*, et al.* (2016). A Common Variant at the 14q32 Endometrial Cancer Risk Locus Activates AKT1 through YY1 Binding. Am. J. Hum. Genet. *98,* 1159-1169.

Paraboschi, E.M., Rimoldi, V., Solda, G., Tabaglio, T., Dall'Osso, C., Saba, E., Vigliano, M., Salviati, A., Leone, M., Benedetti, M.D.*, et al.* (2014). Functional variations modulating PRKCA expression and alternative splicing predispose to multiple sclerosis. Hum. Mol. Genet. *23,* 6746-6761.

Patwardhan, R.P., Hiatt, J.B., Witten, D.M., Kim, M.J., Smith, R.P., May, D., Lee, C., Andrie, J.M., Lee, S.I., Cooper, G.M.*, et al.* (2012). Massively parallel functional dissection of mammalian enhancers in vivo. Nat. Biotechnol. *30,* 265-270.

Peters, J.E., Lyons, P.A., Lee, J.C., Richard, A.C., Fortune, M.D., Newcombe, P.J., Richardson, S., and Smith, K.G. (2016). Insight into Genotype-Phenotype Associations through eQTL Mapping in Multiple Cell Types in Health and Immune-Mediated Disease. PLoS Genet. *12,* e1005908.

Peters, O.M., Cabrera, G.T., Tran, H., Gendron, T.F., McKeon, J.E., Metterville, J., Weiss, A., Wightman, N., Salameh, J., Kim, J.*, et al.* (2015). Human C9ORF72 Hexanucleotide Expansion

Reproduces RNA Foci and Dipeptide Repeat Proteins but Not Neurodegeneration in BAC Transgenic Mice. Neuron *88,* 902-909.

Petkau, T.L., and Leavitt, B.R. (2014). Progranulin in neurodegenerative disease. Trends Neurosci. *37,* 388-398.

Pickford, F., Marcus, J., Camargo, L.M., Xiao, Q., Graham, D., Mo, J.R., Burkhardt, M., Kulkarni, V., Crispino, J., Hering, H., and Hutton, M. (2011). Progranulin is a chemoattractant for microglia and stimulates their endocytic activity. Am. J. Pathol. *178,* 284-295.

Polymenidou, M., Lagier-Tourenne, C., Hutt, K.R., Huelga, S.C., Moran, J., Liang, T.Y., Ling, S.C., Sun, E., Wancewicz, E., Mazur, C.*, et al.* (2011). Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. Nat. Neurosci. *14,* 459-468.

Polymeropoulos, M.H., Lavedan, C., Leroy, E., Ide, S.E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., Boyer, R.*, et al.* (1997). Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science *276,* 2045-2047.

Pombo, A., and Dillon, N. (2015). Three-dimensional genome architecture: players and mechanisms. Nat. Rev. Mol. Cell Biol. *16,* 245-257.

Poorkaj, P., Bird, T.D., Wijsman, E., Nemens, E., Garruto, R.M., Anderson, L., Andreadis, A., Wiederholt, W.C., Raskind, M., and Schellenberg, G.D. (1998). Tau is a candidate gene for chromosome 17 frontotemporal dementia. Ann. Neurol. *43,* 815-825.

Premi, E., Formenti, A., Gazzina, S., Archetti, S., Gasparotti, R., Padovani, A., and Borroni, B. (2014). Effect of TMEM106B polymorphism on functional network connectivity in asymptomatic GRN mutation carriers. JAMA Neurol. *71,* 216-221.

Press, B., Feng, Y., Hoflack, B., and Wandinger-Ness, A. (1998). Mutant Rab7 causes the accumulation of cathepsin D and cation-independent mannose 6-phosphate receptor in an early endocytic compartment. J. Cell Biol. *140,* 1075-1089.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470,* 279-283.

Rademakers, R., Eriksen, J.L., Baker, M., Robinson, T., Ahmed, Z., Lincoln, S.J., Finch, N., Rutherford, N.J., Crook, R.J., Josephs, K.A.*, et al.* (2008). Common variation in the miR-659 binding-site of GRN is a major risk factor for TDP43-positive frontotemporal dementia. Hum. Mol. Genet. *17,* 3631-3642.

Rademakers, R., Neumann, M., and Mackenzie, I.R. (2012). Advances in understanding the molecular basis of frontotemporal dementia. Nat. Rev. Neurol. *8,* 423-434.

Raitano, S., Ordovas, L., De Muynck, L., Guo, W., Espuny-Camacho, I., Geraerts, M., Khurana, S., Vanuytsel, K., Toth, B.I., Voets, T.*, et al.* (2015). Restoration of progranulin expression rescues cortical neuron generation in an induced pluripotent stem cell model of frontotemporal dementia. Stem Cell. Reports *4,* 16-24.

Raj, T., Rothamel, K., Mostafavi, S., Ye, C., Lee, M.N., Replogle, J.M., Feng, T., Lee, M., Asinovski, N., Frohlich, I.*, et al.* (2014). Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. Science *344,* 519-523.

Rajagopal, N., Srinivasan, S., Kooshesh, K., Guo, Y., Edwards, M.D., Banerjee, B., Syed, T., Emons, B.J., Gifford, D.K., and Sherwood, R.I. (2016). High-throughput mapping of regulatory DNA. Nat. Biotechnol. *34,* 167-174.

Ramanan, V.K., and Saykin, A.J. (2013). Pathways to neurodegeneration: mechanistic insights from GWAS in Alzheimer's disease, Parkinson's disease, and related disorders. Am. J. Neurodegener Dis. *2,* 145-175.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159,* 1665-1680.

Rascovsky, K., Hodges, J.R., Knopman, D., Mendez, M.F., Kramer, J.H., Neuhaus, J., van Swieten, J.C., Seelaar, H., Dopper, E.G., Onyike, C.U.*, et al.* (2011). Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. Brain *134,* 2456-2477.

Ratnavalli, E., Brayne, C., Dawson, K., and Hodges, J.R. (2002). The prevalence of frontotemporal dementia. Neurology *58,* 1615-1621.

Ratti, A., and Buratti, E. (2016). Physiological functions and pathobiology of TDP-43 and FUS/TLS proteins. J. Neurochem. *138 Suppl 1,* 95-111.

Remenyi, J., Hunter, C.J., Cole, C., Ando, H., Impey, S., Monk, C.E., Martin, K.J., Barton, G.J., Hutvagner, G., and Arthur, J.S. (2010). Regulation of the miR-212/132 locus by MSK1 and CREB in response to neurotrophins. Biochem. J. *428,* 281-291.

Renton, A.E., Majounie, E., Waite, A., Simon-Sanchez, J., Rollinson, S., Gibbs, J.R., Schymick, J.C., Laaksovirta, H., van Swieten, J.C., Myllykangas, L.*, et al.* (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron *72,* 257-268.

Richardson, K., Nettleton, J.A., Rotllan, N., Tanaka, T., Smith, C.E., Lai, C.Q., Parnell, L.D., Lee, Y.C., Lahti, J., Lemaitre, R.N.*, et al.* (2013). Gain-of-function lipoprotein lipase variant rs13702 modulates lipid traits through disruption of a microRNA-410 seed site. Am. J. Hum. Genet. *92,* 5-14.

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J.*, et al.* (2015). Integrative analysis of 111 reference human epigenomes. Nature *518,* 317-330.

Rollinson, S., Mead, S., Snowden, J., Richardson, A., Rohrer, J., Halliwell, N., Usher, S., Neary, D., Mann, D., Hardy, J., and Pickering-Brown, S. (2011a). Frontotemporal lobar degeneration genome wide association study replication confirms a risk locus shared with amyotrophic lateral sclerosis. Neurobiol. Aging *32,* 758.e1-758.e7.

Rollinson, S., Rohrer, J.D., van der Zee, J., Sleegers, K., Mead, S., Engelborghs, S., Collinge, J., De Deyn, P.P., Mann, D.M., Van Broeckhoven, C., and Pickering-Brown, S.M. (2011b). No

association of PGRN 3'UTR rs5848 in frontotemporal lobar degeneration. Neurobiol. Aging *32,* 754-755.

Rosen, E.Y., Wexler, E.M., Versano, R., Coppola, G., Gao, F., Winden, K.D., Oldham, M.C., Martens, L.H., Zhou, P., Farese, R.V.,Jr, and Geschwind, D.H. (2011). Functional genomic analyses identify pathways dysregulated by progranulin deficiency, implicating Wnt signaling. Neuron *71,* 1030-1042.

Russ, J., Liu, E.Y., Wu, K., Neal, D., Suh, E., Irwin, D.J., McMillan, C.T., Harms, M.B., Cairns, N.J., Wood, E.M.*, et al.* (2015). Hypermethylation of repeat expanded C9orf72 is a clinical and molecular disease modifier. Acta Neuropathol. *129,* 39-52.

Rutherford, N.J., Carrasquillo, M.M., Li, M., Bisceglio, G., Menke, J., Josephs, K.A., Parisi, J.E., Petersen, R.C., Graff-Radford, N.R., Younkin, S.G., Dickson, D.W., and Rademakers, R. (2012). TMEM106B risk variant is implicated in the pathologic presentation of Alzheimer disease. Neurology *79,* 717-718.

Sanjana, N.E., Wright, J., Zheng, K., Shalem, O., Fontanillas, P., Joung, J., Cheng, C., Regev, A., and Zhang, F. (2016). High-resolution interrogation of functional elements in the noncoding genome. Science *353,* 1545-1549.

Sankaran, V.G., Lettre, G., Orkin, S.H., and Hirschhorn, J.N. (2010). Modifier genes in Mendelian disorders: the example of hemoglobin disorders. Ann. N. Y. Acad. Sci. *1214,* 47-56.

Sardiello, M., Palmieri, M., di Ronza, A., Medina, D.L., Valenza, M., Gennarino, V.A., Di Malta, C., Donaudy, F., Embrione, V., Polishchuk, R.S.*, et al.* (2009). A gene network regulating lysosomal biogenesis and function. Science *325,* 473-477.

Sareen, D., O'Rourke, J.G., Meera, P., Muhammad, A.K., Grant, S., Simpkinson, M., Bell, S., Carmona, S., Ornelas, L., Sahabian, A.*, et al.* (2013). Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9ORF72 repeat expansion. Sci. Transl. Med. *5,* 208ra149.

Schaub, M.A., Boyle, A.P., Kundaje, A., Batzoglou, S., and Snyder, M. (2012). Linking disease associations with regulatory information in the human genome. Genome Res. *22,* 1748-1759.

Scheltens, P., Blennow, K., Breteler, M.M., de Strooper, B., Frisoni, G.B., Salloway, S., and Van der Flier, W.M. (2016). Alzheimer's disease. Lancet *388,* 505-517.

Schenk, M.F., Szendro, I.G., Salverda, M.L., Krug, J., and de Visser, J.A. (2013). Patterns of Epistasis between Beneficial Mutations in an Antibiotic Resistance Gene. Mol. Biol. Evol.

Schmidt, E.M., Zhang, J., Zhou, W., Chen, J., Mohlke, K.L., Chen, Y.E., and Willer, C.J. (2015). GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. Bioinformatics *31,* 2601-2606.

Schwenk, B.M., Lang, C.M., Hogl, S., Tahirovic, S., Orozco, D., Rentzsch, K., Lichtenthaler, S.F., Hoogenraad, C.C., Capell, A., Haass, C., and Edbauer, D. (2014). The FTLD risk factor TMEM106B and MAP6 control dendritic trafficking of lysosomes. Embo J. *33,* 450-467.

Scott, H.L., Tamagnini, F., Narduzzo, K.E., Howarth, J.L., Lee, Y.B., Wong, L.F., Brown, M.W., Warburton, E.C., Bashir, Z.I., and Uney, J.B. (2012). MicroRNA-132 regulates recognition memory and synaptic plasticity in the perirhinal cortex. Eur. J. Neurosci. *36,* 2941-2948.

Seelaar, H., Rohrer, J.D., Pijnenburg, Y.A., Fox, N.C., and van Swieten, J.C. (2011). Clinical, genetic and pathological heterogeneity of frontotemporal dementia: a review. J. Neurol. Neurosurg. Psychiatry. *82,* 476-486.

Sekar, A., Bialas, A.R., de Rivera, H., Davis, A., Hammond, T.R., Kamitaki, N., Tooley, K., Presumey, J., Baum, M., Van Doren, V.*, et al.* (2016). Schizophrenia risk from complex variation of complement component 4. Nature *530,* 177-183.

Sellier, C., Campanari, M.L., Julie Corbier, C., Gaucherot, A., Kolb-Cheynel, I., Oulad-Abdelghani, M., Ruffenach, F., Page, A., Ciura, S., Kabashi, E., and Charlet-Berguerand, N. (2016). Loss of C9ORF72 impairs autophagy and synergizes with polyQ Ataxin-2 to induce motor neuron dysfunction and cell death. Embo J. *35,* 1276-1297.

Shankaran, S.S., Capell, A., Hruscha, A.T., Fellerer, K., Neumann, M., Schmid, B., and Haass, C. (2008). Missense mutations in the progranulin gene linked to frontotemporal lobar degeneration with ubiquitin-immunoreactive inclusions reduce progranulin production and secretion. J. Biol. Chem. *283,* 1744-1753.

Shen, S.Q., Myers, C.A., Hughes, A.E., Byrne, L.C., Flannery, J.G., and Corbo, J.C. (2016). Massively parallel cis-regulatory analysis in the mammalian central nervous system. Genome Res. *26,* 238-255.

Sheng, J., Su, L., Xu, Z., and Chen, G. (2014). Progranulin polymorphism rs5848 is associated with increased risk of Alzheimer's disease. Gene *542,* 141-145.

Shi, J., Wang, E., Milazzo, J.P., Wang, Z., Kinney, J.B., and Vakoc, C.R. (2015). Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. Nat. Biotechnol. *33,* 661-667.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nat. Rev. Genet. *15,* 272-286.

Silva, R.F., Mendonca, S.C., Carvalho, L.M., Reis, A.M., Gordo, I., Trindade, S., and Dionisio, F. (2011). Pervasive sign epistasis between conjugative plasmids and drug-resistance chromosomal mutations. PLoS Genet. *7,* e1002181.

Simon-Sanchez, J., Seelaar, H., Bochdanovits, Z., Deeg, D.J., van Swieten, J.C., and Heutink, P. (2009). Variation at GRN 3'-UTR rs5848 is not associated with a risk of frontotemporal lobar degeneration in Dutch population. PLoS One *4,* e7494.

Singleton, A.B., Farrer, M., Johnson, J., Singleton, A., Hague, S., Kachergus, J., Hulihan, M., Peuralinna, T., Dutra, A., Nussbaum, R.*, et al.* (2003). alpha-Synuclein locus triplication causes Parkinson's disease. Science *302,* 841.

Skibinski, G., Parkinson, N.J., Brown, J.M., Chakrabarti, L., Lloyd, S.L., Hummerich, H., Nielsen, J.E., Hodges, J.R., Spillantini, M.G., Thusgaard, T.*, et al.* (2005). Mutations in the endosomal ESCRTIII-complex subunit CHMP2B in frontotemporal dementia. Nat. Genet. *37,* 806-808.

Smemo, S., Tena, J.J., Kim, K.H., Gamazon, E.R., Sakabe, N.J., Gomez-Marin, C., Aneas, I., Credidio, F.L., Sobreira, D.R., Wasserman, N.F., *et al.* (2014). Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature *507,* 371-375.

Smith, K.R., Damiano, J., Franceschetti, S., Carpenter, S., Canafoglia, L., Morbin, M., Rossi, G., Pareyson, D., Mole, S.E., Staropoli, J.F., *et al.* (2012). Strikingly different clinicopathological phenotypes determined by progranulin-mutation dosage. Am. J. Hum. Genet. *90,* 1102-1107.

Smith, R.P., Taher, L., Patwardhan, R.P., Kim, M.J., Inoue, F., Shendure, J., Ovcharenko, I., and Ahituv, N. (2013). Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. Nat. Genet. *45,* 1021-1028.

Snowden, J.S., Rollinson, S., Thompson, J.C., Harris, J.M., Stopford, C.L., Richardson, A.M., Jones, M., Gerhard, A., Davidson, Y.S., Robinson, A., *et al.* (2012). Distinct clinical and pathological characteristics of frontotemporal dementia associated with C9ORF72 mutations. Brain *135,* 693-708.

Soldner, F., Stelzer, Y., Shivalila, C.S., Abraham, B.J., Latourelle, J.C., Barrasa, M.I., Goldmann, J., Myers, R.H., Young, R.A., and Jaenisch, R. (2016). Parkinson-associated risk variant in distal enhancer of alpha-synuclein modulates target gene expression. Nature *533,* 95-99.

Spillantini, M.G., Crowther, R.A., Kamphorst, W., Heutink, P., and van Swieten, J.C. (1998). Tau pathology in two Dutch families with mutations in the microtubule-binding region of tau. Am. J. Pathol. *153,* 1359-1363.

Spillantini, M.G., Schmidt, M.L., Lee, V.M., Trojanowski, J.Q., Jakes, R., and Goedert, M. (1997). Alpha-synuclein in Lewy bodies. Nature *388,* 839-840.

Spisak, S., Lawrenson, K., Fu, Y., Csabai, I., Cottman, R.T., Seo, J.H., Haiman, C., Han, Y., Lenci, R., Li, Q., *et al.* (2015). CAUSEL: an epigenome- and genome-editing pipeline for establishing function of noncoding GWAS variants. Nat. Med. *21,* 1357-1363.

Stadhouders, R., Aktuna, S., Thongjuea, S., Aghajanirefah, A., Pourfarzad, F., van Ijcken, W., Lenhard, B., Rooks, H., Best, S., Menzel, S., *et al.* (2014). HBS1L-MYB intergenic variants modulate fetal hemoglobin via long-range MYB enhancers. J. Clin. Invest. *124,* 1699-1710.

Stagi, M., Klein, Z.A., Gould, T.J., Bewersdorf, J., and Strittmatter, S.M. (2014). Lysosome size, motility and stress response regulated by fronto-temporal dementia modifier TMEM106B. Mol. Cell. Neurosci. *61,* 226-240.

Stead, J.A., Keen, J.N., and McDowall, K.J. (2006). The identification of nucleic acid-interacting proteins using a simple proteomics-based approach that directly incorporates the electrophoretic mobility shift assay. Mol. Cell. Proteomics *5,* 1697-1702.

Stewart, H., Rutherford, N.J., Briemberg, H., Krieger, C., Cashman, N., Fabros, M., Baker, M., Fok, A., DeJesus-Hernandez, M., Eisen, A., Rademakers, R., and Mackenzie, I.R. (2012). Clinical and pathological features of amyotrophic lateral sclerosis caused by mutation in the C9ORF72 gene on chromosome 9p. Acta Neuropathol. *123,* 409-417.

Stranger, B.E., Montgomery, S.B., Dimas, A.S., Parts, L., Stegle, O., Ingle, C.E., Sekowska, M., Smith, G.D., Evans, D., Gutierrez-Arcelus, M.*, et al.* (2012). Patterns of cis regulatory variation in diverse human populations. PLoS Genet. *8,* e1002639.

Strong, M.J., Grace, G.M., Freedman, M., Lomen-Hoerth, C., Woolley, S., Goldstein, L.H., Murphy, J., Shoesmith, C., Rosenfeld, J., Leigh, P.N.*, et al.* (2009). Consensus criteria for the diagnosis of frontotemporal cognitive and behavioural syndromes in amyotrophic lateral sclerosis. Amyotroph Lateral Scler. *10,* 131-146.

Sudria-Lopez, E., Koppers, M., de Wit, M., van der Meer, C., Westeneng, H.J., Zundel, C.A., Youssef, S.A., Harkema, L., de Bruin, A., Veldink, J.H., van den Berg, L.H., and Pasterkamp, R.J. (2016). Full ablation of C9orf72 in mice causes immune system-related pathology and neoplastic events but no motor neuron defects. Acta Neuropathol. *132,* 145-147.

Sullivan, P.M., Zhou, X., Robins, A.M., Paushter, D.H., Kim, D., Smolka, M.B., and Hu, F. (2016). The ALS/FTLD associated protein C9orf72 associates with SMCR8 and WDR41 to regulate the autophagy-lysosome pathway. Acta Neuropathol. Commun. *4,* 51-016-0324-5.

Suzuki, H., and Matsuoka, M. (2016). The Lysosomal Trafficking Transmembrane Protein 106B Is Linked to Cell Death. J. Biol. Chem. *291,* 21448-21460.

Tak, Y.G., and Farnham, P.J. (2015). Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. Epigenetics Chromatin *8,* 57-015-0050-4. eCollection 2015.

Tan, R.H., Ke, Y.D., Ittner, L.M., and Halliday, G.M. (2017). ALS/FTLD: experimental models and reality. Acta Neuropathol.

Tanaka, Y., Matsuwaki, T., Yamanouchi, K., and Nishihara, M. (2013a). Exacerbated inflammatory responses related to activated microglia after traumatic brain injury in progranulin-deficient mice. Neuroscience *231,* 49-60.

Tanaka, Y., Matsuwaki, T., Yamanouchi, K., and Nishihara, M. (2013b). Increased lysosomal biogenesis in activated microglia and exacerbated neuronal damage after traumatic brain injury in progranulin-deficient mice. Neuroscience *250,* 8-19.

Tanaka, Y., Suzuki, G., Matsuwaki, T., Hosokawa, M., Serrano, G., Beach, T.G., Yamanouchi, K., Hasegawa, M., and Nishihara, M. (2017). Progranulin regulates lysosomal function and biogenesis through acidification of lysosomes. Hum. Mol. Genet.

Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Ruszczycki, B.*, et al.* (2015). CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. Cell *163,* 1611-1627.

Tao, Z., Wang, H., Xia, Q., Li, K., Li, K., Jiang, X., Xu, G., Wang, G., and Ying, Z. (2015). Nucleolar stress and impaired stress granule formation contribute to C9orf72 RAN translation-induced cytotoxicity. Hum. Mol. Genet. *24,* 2426-2441.

Tehranchi, A.K., Myrthil, M., Martin, T., Hie, B.L., Golan, D., and Fraser, H.B. (2016). Pooled ChIP-Seq Links Variation in Transcription Factor Binding to Complex Disease Risk. Cell *165,* 730-741.

Tewhey, R., Kotliar, D., Park, D.S., Liu, B., Winnicki, S., Reilly, S.K., Andersen, K.G., Mikkelsen, T.S., Lander, E.S., Schaffner, S.F., and Sabeti, P.C. (2016). Direct Identification of Hundreds of Expression-Modulating Variants using a Multiplexed Reporter Assay. Cell *165,* 1519-1529.

Therrien, M., Rouleau, G.A., Dion, P.A., and Parker, J.A. (2013). Deletion of C9ORF72 results in motor neuron degeneration and stress sensitivity in C. elegans. PLoS One *8,* e83450.

Todd, T.W., and Petrucelli, L. (2016). Insights into the pathogenic mechanisms of Chromosome 9 open reading frame 72 (C9orf72) repeat expansions. J. Neurochem. *138 Suppl 1,* 145-162.

Toledo, J.B., Van Deerlin, V.M., Lee, E.B., Suh, E., Baek, Y., Robinson, J.L., Xie, S.X., McBride, J., Wood, E.M., Schuck, T.*, et al.* (2013). A platform for discovery: The University of Pennsylvania Integrated Neurodegenerative Disease Biobank. Alzheimers Dement.

Tollervey, J.R., Curk, T., Rogelj, B., Briese, M., Cereda, M., Kayikci, M., Konig, J., Hortobagyi, T., Nishimura, A.L., Zupunski, V.*, et al.* (2011). Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. Nat. Neurosci. *14,* 452-458.

Toulorge, D., Schapira, A.H., and Hajj, R. (2016). Molecular changes in the postmortem parkinsonian brain. J. Neurochem. *139 Suppl 1,* 27-58.

Tran, H., Almeida, S., Moore, J., Gendron, T.F., Chalasani, U., Lu, Y., Du, X., Nickerson, J.A., Petrucelli, L., Weng, Z., and Gao, F.B. (2015). Differential Toxicity of Nuclear RNA Foci versus Dipeptide Repeat Proteins in a Drosophila Model of C9ORF72 FTD/ALS. Neuron *87,* 1207-1214.

Tseng, H.C., Ruegg, S.J., Maronski, M., Messam, C.A., Grinspan, J.B., and Dichter, M.A. (2006). Injuring neurons induces neuronal differentiation in a population of hippocampal precursor cells in culture. Neurobiol. Dis. *22,* 88-97.

Ugolino, J., Ji, Y.J., Conchina, K., Chu, J., Nirujogi, R.S., Pandey, A., Brady, N.R., Hamacher-Brady, A., and Wang, J. (2016). Loss of C9orf72 Enhances Autophagic Activity via Deregulated mTOR and TFEB Signaling. PLoS Genet. *12,* e1006443.

Ulirsch, J.C., Nandakumar, S.K., Wang, L., Giani, F.C., Zhang, X., Rogov, P., Melnikov, A., McDonel, P., Do, R., Mikkelsen, T.S., and Sankaran, V.G. (2016). Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits. Cell *165,* 1530-1545.

van Blitterswijk, M., Mullen, B., Nicholson, A.M., Bieniek, K.F., Heckman, M.G., Baker, M.C., DeJesus-Hernandez, M., Finch, N.A., Brown, P.H., Murray, M.E.*, et al.* (2014). TMEM106B protects C9ORF72 expansion carriers against frontotemporal dementia. Acta Neuropathol. *127,* 397-406.

Van Deerlin, V.M., Sleiman, P.M., Martinez-Lage, M., Chen-Plotkin, A., Wang, L.S., Graff-Radford, N.R., Dickson, D.W., Rademakers, R., Boeve, B.F., Grossman, M.*, et al.* (2010). Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. Nat. Genet. *42,* 234-239.

van der Zee, J., and Van Broeckhoven, C. (2011). TMEM106B a novel risk factor for frontotemporal lobar degeneration. J. Mol. Neurosci. *45,* 516-521.

van der Zee, J., Van Langenhove, T., Kleinberger, G., Sleegers, K., Engelborghs, S., Vandenberghe, R., Santens, P., Van den Broeck, M., Joris, G., Brys, J.*, et al.* (2011). TMEM106B

is associated with frontotemporal lobar degeneration in a clinically diagnosed patient cohort. Brain *134,* 808-815.

Vass, R., Ashbridge, E., Geser, F., Hu, W.T., Grossman, M., Clay-Falcone, D., Elman, L., McCluskey, L., Lee, V.M., Van Deerlin, V.M., Trojanowski, J.Q., and Chen-Plotkin, A.S. (2011). Risk genotypes at TMEM106B are associated with cognitive impairment in amyotrophic lateral sclerosis. Acta Neuropathol. *121,* 373-380.

Verlaan, D.J., Berlivet, S., Hunninghake, G.M., Madore, A.M., Lariviere, M., Moussette, S., Grundberg, E., Kwan, T., Ouimet, M., Ge, B.*, et al.* (2009). Allele-specific chromatin remodeling in the ZPBP2/GSDMB/ORMDL3 locus associated with the risk of asthma and autoimmune disease. Am. J. Hum. Genet. *85,* 377-393.

Vicente, C.T., Edwards, S.L., Hillman, K.M., Kaufmann, S., Mitchell, H., Bain, L., Glubb, D.M., Lee, J.S., French, J.D., and Ferreira, M.A. (2015). Long-Range Modulation of PAG1 Expression by 8q21 Allergy Risk Variants. Am. J. Hum. Genet. *97,* 329-336.

Visser, M., Kayser, M., and Palstra, R.J. (2012). HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. Genome Res. *22,* 446-455.

Vo, N., Klein, M.E., Varlamova, O., Keller, D.M., Yamamoto, T., Goodman, R.H., and Impey, S. (2005). A cAMP-response element binding protein-induced microRNA regulates neuronal morphogenesis. Proc. Natl. Acad. Sci. U. S. A. *102,* 16426-16431.

Vockley, C.M., Guo, C., Majoros, W.H., Nodzenski, M., Scholtens, D.M., Hayes, M.G., Lowe, W.L.,Jr, and Reddy, T.E. (2015). Massively parallel quantification of the regulatory effects of noncoding genetic variation in a human cohort. Genome Res. *25,* 1206-1214.

Wang, D., Poi, M.J., Sun, X., Gaedigk, A., Leeder, J.S., and Sadee, W. (2014a). Common CYP2D6 polymorphisms affecting alternative splicing and transcription: long-range haplotypes with two regulatory variants modulate CYP2D6 activity. Hum. Mol. Genet. *23,* 268-278.

Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014b). Genetic screens in human cells using the CRISPR-Cas9 system. Science *343,* 80-84.

Wang, W.X., Rajeev, B.W., Stromberg, A.J., Ren, N., Tang, G., Huang, Q., Rigoutsos, I., and Nelson, P.T. (2008). The expression of microRNA miR-107 decreases early in Alzheimer's disease and may accelerate disease progression through regulation of beta-site amyloid precursor protein-cleaving enzyme 1. J. Neurosci. *28,* 1213-1223.

Wang, X., Tucker, N.R., Rizki, G., Mills, R., Krijger, P.H., de Wit, E., Subramanian, V., Bartell, E., Nguyen, X.X., Ye, J.*, et al.* (2016). Discovery and validation of sub-threshold genome-wide association study loci using epigenomic signatures. Elife *5,* 10.7554/eLife.10557.

Wang, X.Q., and Dostie, J. (2016). Chromosome folding and its regulation in health and disease. Curr. Opin. Genet. Dev. *43,* 23-30.

Ward, L.D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. *40,* D930-4.

Watts, G.D., Wymer, J., Kovach, M.J., Mehta, S.G., Mumm, S., Darvish, D., Pestronk, A., Whyte, M.P., and Kimonis, V.E. (2004). Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. Nat. Genet. *36,* 377-381.

Wayman, G.A., Davare, M., Ando, H., Fortin, D., Varlamova, O., Cheng, H.Y., Marks, D., Obrietan, K., Soderling, T.R., Goodman, R.H., and Impey, S. (2008). An activity-regulated microRNA controls dendritic plasticity by down-regulating p250GAP. Proc. Natl. Acad. Sci. U. S. A. *105,* 9093-9098.

Webster, C.P., Smith, E.F., Bauer, C.S., Moller, A., Hautbergue, G.M., Ferraiuolo, L., Myszczynska, M.A., Higginbottom, A., Walsh, M.J., Whitworth, A.J.*, et al.* (2016). The C9orf72 protein interacts with Rab1a and the ULK1 complex to regulate initiation of autophagy. Embo J. *35,* 1656-1676.

Weinreich, D.M., Watson, R.A., and Chao, L. (2005). Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. Evolution *59,* 1165-1174.

Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature *447,* 661-678.

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., and Parkinson, H. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. *42,* D1001-6.

Wen, X., Tan, W., Westergard, T., Krishnamurthy, K., Markandaiah, S.S., Shi, Y., Lin, S., Shneider, N.A., Monaghan, J., Pandey, U.B.*, et al.* (2014). Antisense proline-arginine RAN dipeptides linked to C9ORF72-ALS/FTD form toxic nuclear aggregates that initiate in vitro and in vivo neuronal death. Neuron *84,* 1213-1225.

Wettenhall, J.M., and Smyth, G.K. (2004). limmaGUI: a graphical user interface for linear modeling of microarray data. Bioinformatics *20,* 3705-3706.

Wijsman, E.M., Daw, E.W., Yu, X., Steinbart, E.J., Nochlin, D., Bird, T.D., and Schellenberg, G.D. (2005). APOE and other loci affect age-at-onset in Alzheimer's disease families with PS2 mutation. Am. J. Med. Genet. B. Neuropsychiatr. Genet. *132B,* 14-20.

Williams, R.L.,Jr, Starmer, J., Mugford, J.W., Calabrese, J.M., Mieczkowski, P., Yee, D., and Magnuson, T. (2014). fourSig: a method for determining chromosomal interactions in 4C-Seq data. Nucleic Acids Res. *42,* e68.

Wils, H., Kleinberger, G., Pereson, S., Janssens, J., Capell, A., Van Dam, D., Cuijt, I., Joris, G., De Deyn, P.P., Haass, C., Van Broeckhoven, C., and Kumar-Singh, S. (2012). Cellular ageing, increased mortality and FTLD-TDP-associated neuropathology in progranulin knockout mice. J. Pathol. *228,* 67-76.

Won, H., de la Torre-Ubieta, L., Stein, J.L., Parikshak, N.N., Huang, J., Opland, C.K., Gandal, M.J., Sutton, G.J., Hormozdiari, F., Lu, D.*, et al.* (2016). Chromosome conformation elucidates regulatory relationships in developing human brain. Nature *538,* 523-527.

Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z.*, et al.* (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. Nat. Genet. *46,* 1173-1186.

Wu, L., Candille, S.I., Choi, Y., Xie, D., Jiang, L., Li-Pook-Than, J., Tang, H., and Snyder, M. (2013). Variation and genetic control of protein abundance in humans. Nature *499,* 79-82.

Wu, Y., Gao, H., Li, H., Tabara, Y., Nakatochi, M., Chiu, Y.F., Park, E.J., Wen, W., Adair, L.S., Borja, J.B.*, et al.* (2014). A meta-analysis of genome-wide association studies for adiponectin levels in East Asians identifies a novel locus near WDR11-FGFR2. Hum. Mol. Genet. *23,* 1108-1119.

Xie, S.X., Baek, Y., Grossman, M., Arnold, S.E., Karlawish, J., Siderowf, A., Hurtig, H., Elman, L., McCluskey, L., Van Deerlin, V., Lee, V.M., and Trojanowski, J.Q. (2011). Building an integrated neurodegenerative disease database at an academic health center. Alzheimers Dement. *7,* e84-93.

Xu, Z., Poidevin, M., Li, X., Li, Y., Shu, L., Nelson, D.L., Li, H., Hales, C.M., Gearing, M., Wingo, T.S., and Jin, P. (2013). Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. Proc. Natl. Acad. Sci. U. S. A. *110,* 7778-7783.

Yang, D., Abdallah, A., Li, Z., Lu, Y., Almeida, S., and Gao, F.B. (2015). FTD/ALS-associated poly(GR) protein impairs the Notch pathway and is recruited by poly(GA) into cytoplasmic inclusions. Acta Neuropathol. *130,* 525-535.

Yang, M., Liang, C., Swaminathan, K., Herrlinger, S., Lai, F., Shiekhattar, R., and Chen, J.F. (2016). A C9ORF72/SMCR8-containing complex regulates ULK1 and plays a dual role in autophagy. Sci. Adv. *2,* e1601167.

Yin, F., Banerjee, R., Thomas, B., Zhou, P., Qian, L., Jia, T., Ma, X., Ma, Y., Iadecola, C., Beal, M.F., Nathan, C., and Ding, A. (2010a). Exaggerated inflammation, impaired host defense, and neuropathology in progranulin-deficient mice. J. Exp. Med. *207,* 117-128.

Yin, F., Dumont, M., Banerjee, R., Ma, Y., Li, H., Lin, M.T., Beal, M.F., Nathan, C., Thomas, B., and Ding, A. (2010b). Behavioral deficits and progressive neuropathology in progranulin-deficient mice: a mouse model of frontotemporal dementia. Faseb J. *24,* 4639-4647.

Yu, L., De Jager, P.L., Yang, J., Trojanowski, J.Q., Bennett, D.A., and Schneider, J.A. (2015). The TMEM106B locus and TDP-43 pathology in older persons without FTLD. Neurology *84,* 927-934.

Zhang, K., Donnelly, C.J., Haeusler, A.R., Grima, J.C., Machamer, J.B., Steinwald, P., Daley, E.L., Miller, S.J., Cunningham, K.M., Vidensky, S.*, et al.* (2015). The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. Nature *525,* 56-61.

Zhang, L., Liu, Y., Song, F., Zheng, H., Hu, L., Lu, H., Liu, P., Hao, X., Zhang, W., and Chen, K. (2011). Functional SNP in the microRNA-367 binding site in the 3'UTR of the calcium channel ryanodine receptor gene 3 (RYR3) affects breast cancer risk and calcification. Proc. Natl. Acad. Sci. U. S. A. *108,* 13653-13658.

Zhang, W., Dourado, D.F., Fernandes, P.A., Ramos, M.J., and Mannervik, B. (2012). Multidimensional epistasis and fitness landscapes in enzyme evolution. Biochem. J. *445,* 39-46.

Zhang, Y.J., Gendron, T.F., Grima, J.C., Sasaguri, H., Jansen-West, K., Xu, Y.F., Katzman, R.B., Gass, J., Murray, M.E., Shinohara, M.*, et al.* (2016). C9ORF72 poly(GA) aggregates sequester and impair HR23 and nucleocytoplasmic transport proteins. Nat. Neurosci. *19,* 668-677.

Zheng, J., Huang, X., Tan, W., Yu, D., Du, Z., Chang, J., Wei, L., Han, Y., Wang, C., Che, X.*, et al.* (2016). Pancreatic cancer risk variant in LINC00673 creates a miR-1231 binding site and interferes with PTPN11 degradation. Nat. Genet. *48,* 747-757.

Zhou, X., Baron, R.M., Hardin, M., Cho, M.H., Zielinski, J., Hawrylkiewicz, I., Sliwinski, P., Hersh, C.P., Mancini, J.D., Lu, K.*, et al.* (2012). Identification of a chronic obstructive pulmonary disease genetic determinant that regulates HHIP. Hum. Mol. Genet. *21,* 1325-1335.

Zhou, X., and Wang, T. (2012). Using the Wash U Epigenome Browser to examine genome-wide sequencing data. Curr. Protoc. Bioinformatics *Chapter 10,* Unit10.10.

Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., and Yang, J. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet. *48,* 481-487.

Zu, T., Gibbens, B., Doty, N.S., Gomes-Pereira, M., Huguet, A., Stone, M.D., Margolis, J., Peterson, M., Markowski, T.W., Ingram, M.A.*, et al.* (2011). Non-ATG-initiated translation directed by microsatellite expansions. Proc. Natl. Acad. Sci. U. S. A. *108,* 260-265.

Zu, T., Liu, Y., Banez-Coronel, M., Reid, T., Pletnikova, O., Lewis, J., Miller, T.M., Harms, M.B., Falchook, A.E., Subramony, S.H.*, et al.* (2013). RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. Proc. Natl. Acad. Sci. U. S. A. *110,* E4968-77.

Zuin, J., Dixon, J.R., van der Reijden, M.I., Ye, Z., Kolovos, P., Brouwer, R.W., van de Corput, M.P., van de Werken, H.J., Knoch, T.A., van IJcken, W.F.*, et al.* (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc. Natl. Acad. Sci. U. S. A. *111,* 996-1001.